Master of Science Thesis in Electrical Engineering
Department of Electrical Engineering, Linköping University, 2022

Investigating Tweet Propagation via Dynamical Models and Influencer Analysis

Joel Nilsson

LINKÖPING UNIVERSITY

Master of Science Thesis in Electrical Engineering

Investigating Tweet Propagation via Dynamical Models and Influencer Analysis

Joel Nilsson

LiTH-ISY-EX--22/5483--SE

Supervisor: **Alberto Zenere**
ISY, Linköpings universitet

Examiner: **Claudio Altafini**
ISY, Linköpings universitet

Division of Automatic Control
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Abstract

Social media consume an increasing portion of people's daily lives and are important platforms in the realms of politics and marketing for reaching out to voters and consumers. Describing and predicting the behaviour of users on social media is thus of interest for companies and politicians, as well as researchers studying information diffusion and human behaviour.

Twitter is a fast-paced microblog that is host to debates, conversations, and campaigns between users as well as organisations all over the world. As all interactions on Twitter are publicly available, the platform has been used as a data source for many studies. While previous works have mainly focused on interaction dynamics for specific user groups or topics, or on predicting virality, the perspective we take in this thesis is to focus on the level of the individual conversation and to use dynamical models to characterise user interactions.

The most prominent characteristic of Twitter conversations is the clear presence of peaks in engagement. We introduce a classification scheme based on peak configurations to quantify the interaction patterns present on Twitter and find that around 70% of conversations exhibit a single peak in user engagement, usually followed by a slower decay. A second order linear model describes the dynamics of the single peak scenario well, indicating that most conversations have two phases - an initial phase of rapid rise and decline in interaction rate, followed by a phase of slowly decreasing interaction rate. We quantify the characteristic life span of Twitter conversations in terms of the second order system time constants.

Furthermore, we investigate the impact that users with many followers, so called *influencers*, have on conversation dynamics, and in particular on the emergence of interaction peaks. The data suggests that influencers do have a noticeable, albeit limited effect on the spreading of conversations to other users.

Acknowledgments

Working on this thesis has been fun and a great learning experience. I would like to thank Prof. Claudio Altafini for providing me the opportunity to write a thesis on this topic, and for taking a highly active role in guiding me throughout the process. I would also like to thank Alberto Zenere for constructive feedback and excellent writing advice.

I would like to thank Oscar Hermansson and Marco Ratta for being wonderful company at the office, and Emil Eriksson for thorough feedback during the opposition. Finally, I am truly grateful to my friends and family for supporting me all these years.

Linköping, June 2022
Joel Nilsson

Contents

Notation

Mathematical notation

Notation	Meaning
$\mathbb{N}$	The set of natural numbers
$\mathbb{R}$	The set of real numbers
$[a, b)$	The half-open interval $\{x \in \mathbb{R} \mid a \leq x < b\}$
$H(t)$	The Heaviside function
$\delta(t)$	The Dirac delta distribution

Abbreviations

Abbreviation	Meaning
AIC	Akaike information criterion
API	Application programming interface
JSON	Javascript object notation
RSS	Residual sum of squares
UTC	Coordinated Universal Time

1 Introduction

The increasing number of people with access to the Internet is making the world a more connected place. Humans are social creatures and spend a significant portion of their days using social media, not only as a means of socialising, but also sharing ideas and consuming news. The rapid velocity at which information is produced and made available can have a positive impact on society. It also changes the way in which humans interact and pay attention to topics. Findings in [11] indicate that the rise and decay rates of collective attention to cultural products or topics are increasing, a phenomenon that is not restricted to social media. The limited amount of collective attention along with an increase in content production shortens attention cycles. Twitter is a microblogging platform where these phenomena can be observed. In [17], cohorts of users on Twitter were clustered into different types depending on the frequency of their activities. The authors found that users who joined Twitter later tend to be more active, accelerating the pace of interactions and exhausting topics more quickly. Since Twitter's founding in 2006 it has seen tremendous growth, having 229M daily active users[1] that generate around 500M tweets per day[2]. Twitter is host to discussions of a plethora of topics, and has become a popular object of study due to the availability of data through the Twitter API. Social network analysis is the study of relationships and interactions between groups or individuals. The field has gained a lot of attention following the development of tools for analysing social networks, notably the theory on complex networks.

Complex networks arise naturally in many settings and are characterised by "a large number of units interconnected through highly non-trivial patterns of interactions" [8]. Examples include the neurons in the human brain, the webpages on the World Wide Web, and humans interacting with each other. Real world

[1] investor.twitterinc.com
[2] internetlivestats.com

networks commonly display complex structures that are distinct from random networks [2]. This is true for social networks as well, in fact, human behaviour is characterised by heavy tailed distributions, both in terms of who knows whom in a social network [2], and the distribution of interarrival times of human activities [1]. Networks can change over time; in the case of social networks, this happens when people make new acquaintances or lose contact with old friends. Twitter is a good example of this: the users in the growing Twitter network are constantly following and unfollowing other users at a substantial rate, a dynamic that is at times accelerated by information cascades [12]. One can also consider conversations on Twitter as graphs, with users as nodes and the links representing interactions.

1.1 Motivation

Previous work on social media, and Twitter in particular, focuses mainly on analysing diffusion of and competition between widely discussed topics [14], predicting the virality of posts [19], and on the marketing potential in online social networks. In [14], the diffusion of hashtags is described with a deterministic model based on epidemiology. The idea is that information on Twitter spreads on a large scale similar to an epidemic, and their model uses constant and time-varying infection rates to capture different spreading behaviours. The authors in [19] proposed a doubly stochastic process to predict the final number of shares of a post. In their model, tweets are retweeted with a time-varying rate that depends on the number of shares, the number of followers of the users who share the post, and a stochastic process that captures engagement received. Studies have also been conducted on the structure of conversations, using branching processes to describe the evolution of conversations [7]. However, so far there have been few investigations on the conversation-level dynamic characteristics of social media interactions, such as the life length of a post. It is also common to use data from a specific domain, e.g., politics, which might be unrepresentative of general behaviour on Twitter.

The purpose of the study is to quantify the temporal characteristics of tweets in order to better understand conversation dynamics on Twitter. In contrast to earlier work, the focus of this study is not directed towards any specific topic or user, but rather on the properties of interactions on Twitter. We investigate the decay of engagement with tweets drawing on linear system theory. The half-life period of a quantity refers to the time within which the concentration is halved on average. The canonical example is radioactive substances, which decay at a rate proportional to the remaining concentration and can be described by a linear differential equation [13]. In [4], the collective attention of humans is modelled in terms of communicative and cultural memory, that decay as new cultural products are generated and outcompete older ones. Half-life has also been used to describe content life-length on social media [18]. We aim to build a model of engagement on Twitter that will give insights into how information spreads on social media, as well as the nature of human interactions online.

1.2 Aim

The goal of this thesis is to characterise the dynamics of Twitter conversations. To do this, we retrieve a dataset sampled from the whole of Twitter, not limited to specific communities or topics. From an initial inspection of interaction patterns, it is evident that tweets commonly arrive in bursts that manifest as peaks in the interaction time series. We aim to develop a method for classifying conversation types based on the pattern of peaks. The classification result should give a hint as to which models are suitable for modelling Twitter dynamics.

Based on the patterns extracted, we investigate how well interactions on Twitter can be described by linear models of the first and second order, and the implications this has for conversation dynamics. We evaluate the pace of conversations by looking at the distribution of time constants of the linear systems. For instance, if a first order model has a good fit, it can be justified to speak of a half-life for conversations. Not all conversations can be modelled by the proposed systems, and we explore other factors that could account for occurrences of peaks in the data. In particular, we analyse the impact of users with a large follower count to provide a heuristic explanation for interaction peaks.

1.3 Delimitations

This work uses linear, deterministic models to describe the interaction patterns on Twitter. We do not consider the content of posts, nor any attributes of the engaging users except follower count when analysing conversation patters.

We are limited in the amount of data we can collect from Twitter. This mainly concerns the user network, which cannot be retrieved at a scale sufficient to compute reliable statistics.

1.4 Thesis Outline

The thesis is structured in the following way. Chapter 2 describes Twitter, the data collected from the platform, and the methods used in this work. Two linear models are proposed to model interactions between users. In Chapter 3, we present our analysis of the data and the results of fitting the models from Chapter 2. We also investigate the impact of users with a high follower count on the arrival of interactions in threads on Twitter. Finally, we discuss and summarise our findings, and outline future work on the topic in Chapter 4.

2

Method

This chapter gives an introduction to Twitter, and goes into detail on the dataset of this study. We introduce some terminology related to tweets and conversations that is used in the analysis, as well as the functions and limitations of the data collection tools. We present a categorisation of conversations useful for describing different interaction patterns, and how to classify conversations accordingly. We introduce models describing the engagement in the time domain, along with the methods we use to fit the models to data and evaluate the results.

2.1 Twitter Basics

Twitter is an online social media network, where individuals, organisations, and institutions can create accounts and post content, also called tweets. A tweet is a single post that may contain some text, limited to 280 (originally 140) characters, and optionally some attached media, e.g., a photo, a movie clip, or a poll. The basic content generating mechanisms on Twitter are: posting a new message, *replying* to a tweet, and sharing a tweet in a new message (*quote*). In general, any of these three actions can be referred to as *tweeting*. A very popular sharing feature of Twitter is called *retweet*, which is used simply to repost a tweet without adding any more content to the post. In general, any tweet, including replies and quotes, may be retweeted, quoted, or replied to. Users may also *like* tweets.

A user can follow other users they find interesting, thus taking part of the content they generate. A user has two personal feeds where tweets from the users they follow appear, governed by Twitter's algorithm. The *Home* feed shows content that Twitter believes the user will care about based on which tweets and users they have interacted with before. This may include content from accounts that the user does not follow. The *Latest Tweets* feed shows posts from all followings in reverse chronological order. This feed is also filtered by Twitter to avoid users getting swamped with content. Users can also create custom feeds by categorising accounts in lists. In addition to these feeds, users can navigate current (personally tailored) trends and topics, or visit a specific user page, all of this through the search functions that Twitter provides.

When using Twitter, one will observe that the Home and Latest Tweets feeds consist mostly of original posts, quotes and retweets. Replies can appear as well, but the impression that the author of this report has at the time of writing is that they are generally more rare. Thus, it seems not all interactions on Twitter are as effective when it comes to propagating a tweet. After posting, retweeting, or quoting a tweet, it will make the tweet likely to appear in the feed of the followers of the user, and it will be visible on the (default) *Tweets* tab of a user's timeline. Replies from a user are visible on the timeline only if one views the *Tweets & replies* tab of the timeline.

A retweet cannot be interacted with in the same way as a tweet, as it is a repost of an original tweet. That is, if a user retweets a post, it will be visible as a tweet from the original author and not by the retweeter, although an indication that the post was retweeted will be visible. A like or a reply to a retweet will count as an interaction with the original tweet. A quote tweet is essentially a retweet but with the addition of new content, which makes it a stand-alone tweet in practice. The quote starts a new thread while providing easy access to the original tweet in the same way a retweet does. A quote can also be a reply, that is, it is possible to quote another tweet when replying in a thread. Liking a tweet will not create a new post, however, it might influence which tweets the Twitter algorithm suggests to the user's followers in their Home feed. Replies and quotes are active ways of engaging with the topic, while retweets and likes can be seen as more passive, since they require minimal effort from the user.

Tweets can contain hashtags and mentions. Hashtags are indicated by the '#'-symbol and they are a way of marking a tweet with a topic (e.g., "#twitter"), which may be picked up and propagated by others. A mention is an '@'-symbol followed by a username (e.g., "@jack"), which sends a notification to that person as an indication that they have received a reply or are encouraged to participate in the conversation. Hashtags and mentions are not considered in this study.

Twitter is arguably more of a public platform in comparison to Facebook in terms of accessibility of the platform content. Most Twitter users have public accounts that can be followed without active consent (or reciprocation), and tweets are by default visible on their timelines. Some users require aspiring followers to send a request in order to gain access to the tweets, in which case the tweets are not accessible from the API either. Users may limit who can reply to a tweet (e.g., only users who are mentioned in the conversation), but this does not affect visibility. If a tweet or account is deleted, the information cannot be accessed.

2.2 Terminology

In this analysis we will look at conversations on Twitter branching out from posts that we call *root tweets* (or simply *roots*). A root tweet must not be a reply to another tweet, nor can it be a retweet, however, root tweets can be quotes of other tweets. We define a *conversation* related to a root tweet to consist of the main interactions towards the tweet. Specifically, we include all retweets and quotes of a root tweet, the root tweet itself, and any tweet in a reply chain connected to the root to be part of its conversation. The retweets and quotes of replies in the conversation are not included in the set of conversation tweets. Unless otherwise stated, by a reply we mean any reply in the conversation, and not necessarily a reply directly towards the root tweet. We will use the words *engagement* and *interaction* interchangeably to mean a retweet, quote, or reply to a conversation.

2.3 The Twitter API

Every day, a massive amount of tweets are produced by Twitter's users, and collecting all of them over a longer period of time is infeasible for reasons discussed later in this section. Twitter provides access to the tweets and users on the platform through an application programming interface (API) which only requires a Twitter account with developer status to use. The API can be queried for tweets in real-time or from the Twitter archive, and may also be used for posting content in an automated manner. There exist many tools for retrieving data through the API. In this study, the command line tool and Python library Twarc[1] is used to fetch data via the Twitter API v2, the newest version of the Twitter API at the time of writing. Twarc automatically queries the API for all relevant data, and structures the retrieved information (payload) in JSON format.

A tweet payload consists of many different fields of data and metadata such as the tweet text, its author, how many likes or retweets it has, the conversation thread it is part of, as well as categorisation of the contents of the tweet (see Table A.5), and much more. A complete list of attributes of Twitter users and tweets can be found in Tables A.3 and A.4 respectively. Information is accessed using a set of so-called endpoints in the API. We present a few of the available endpoints in Table A.6. For instance, tweets can be retrieved based on attributes such as words, or by users that have liked or retweeted a post. User objects can be retrieved through queries on username and user ID, and it is possible to extract follower and following lists of users. Other retrievable objects include Spaces (live audio conversations on Twitter), lists (a shareable list of accounts that can be used as a filter for a user feed), and media objects (polls, places, and other multimedia content).

The functions in the Twarc library retrieve the maximum number of attributes for all objects by default. The `referenced_tweets` attribute may contain a full body of information on a referenced tweet, including author and entity objects, which potentially makes a single tweet payload quite large. It is possible to filter out the most relevant fields when querying from both command line and the Twarc library. Some attributes, such as `id`, are always included in the payload, but not all returned tweet objects contain the same attributes.

[1] twarc-project.readthedocs.io

Table 2.1: Example of query options of the search tweets endpoint.

Query operation	Description
`keyword`	Match a keyword of a tweet
`"exact phrase match"`	Match an exact phrase of a tweet
`retweets_of:X`	Match a retweet of a tweet from user "`X`"
`conversation_id:Y`	Match tweets with conversation ID "`Y`"

For instance, a tweet that does not refer to any other tweet will not contain the `referenced_tweets` attribute.

The *full archive search* endpoint allows us to search for any tweets posted since the dawn of Twitter, filtering by multiple criteria. A few examples are shown in Table 2.1: whether the tweet contains a specific keyword or an exact phrase, if it has a specific conversation ID, or if the tweet in question is a retweet of a tweet from a certain author. Querying on the conversation ID yields the replies of a conversation, and querying on the tweet phrase and providing the author ID will return the retweets of that tweet. The *quote tweets lookup* endpoint is used to retrieve quotes of a given tweet.

The *sample stream* endpoint offers a way to sample tweets in real time. According to Twitter, the sample stream endpoint provides access to a statistically relevant sample of about 1% of all tweets (including retweets, quotes, and original posts). It is not clearly stated what Twitter means by a "statistically relevant sample". If sampling is performed uniformly at random from all types of tweets during some unit of time, there is a higher probability of fetching a reply or retweet related to a popular conversation ID. If this is true, we will sample tweets related to more popular conversations on average. As is explained in the next section, we exclude conversations with fewer than 50 interactions from our dataset, so it matters less if we do not sample the conversations less popular conversations.

Twitter limits the number of requests and the amount of data that can be returned with each request. A basic development project on Twitter may retrieve 500,000 tweets per month, while the Academic access track offers 10M tweets per month. There is a rolling limit on requests that can be sent to the API, called rate limits. Some endpoints have higher limits, such as tweet retrieval with 900 requests in a 15-minute period, while others have low limits, such as user follower lookup which offers 15 requests in a 15-minute period. The restriction on user following lookups puts severe constraints on our ability to construct the social network around users with many followers.

2.4 Data

The dataset used in this work consists of 38,042,280 tweets from 9,758 conversations collected over a period from January to May, in 2022. The sample stream endpoint is used to collect a set of tweets: root tweets, replies, retweets and quotes. From these tweets we collect the conversation IDs of the sampled tweets,

Table 2.2: *Platform used by users in sample of our dataset.*

Platform	#replies
Twitter for Android	1,409,849
Twitter Web App	717,739
Twitter for iPhone	525,831
TweetDeck	36,292
Twitter for iPad	34,335
Twitter Web Client	3,605

or, if the tweet is a retweet, we use the conversation ID of the retweeted post. We then proceed to collect the conversation replies and retweets and quotes for the conversation root. Replies and retweets are collected via the full archive search endpoint, and quotes through the quote tweets lookup endpoint. In order to avoid left-censoring, i.e., missing out on data due to retrieving the conversation when users are still interacting, we wait at least one week from sampling before collecting the conversation IDs and the rest of the interactions. For each tweet we collect a subset of the available attributes, the most important being the time of posting, the tweets to which it makes a reference, and the number of followers of the user posting.

All tweet payloads retrieved via the API optionally include statistics on how many retweets, quotes, replies, and likes the tweet in question has acquired in the `public_metrics` attribute. We often find that there is a difference in the number of tweets we retrieve and the number given by Twitter. We describe the reason behind the difference in more detail in Section A.1. Comparing replies, retweets and quotes, the error is largest for retweets. Prior to the analysis, we filter out the conversations that contain fewer than 50% of the retweets given in the `public_metrics` attribute of the root tweet. We can also collect information on which application was used to post a tweet. Table 2.2 shows the applications from which a subset of the tweets in the dataset were posted. The tweets in the table include 99.59% of all tweets in the subsample, and indicates that third party platforms, that might show tweets differently from the description given in Section 2.1, are uncommon.

2.4.1 Characterising Twitter Conversations

To quantify the engagement of the sampled conversation over time, we divide the time axis into 1-hour wide bins and compute a histogram wherein replies, retweets, and quotes are equally weighted. A resolution of one hour was found to make the histogram sufficiently dense without losing information of when interactions are posted. An example of such a histogram is shown in Figure 2.1. Conversations with few interactions exhibit spurious patterns and peaks which adds uncertainty to the statistics in the analysis. Therefore, we only consider conversations with at least 50 interactions.

Interaction patterns in Twitter conversations take many different shapes and

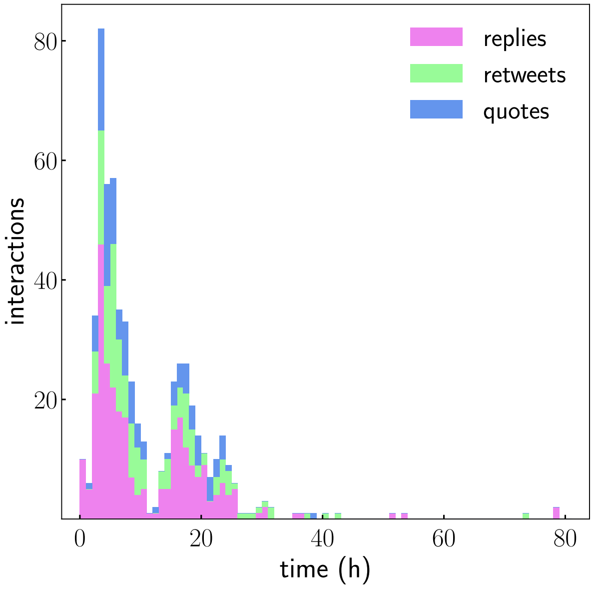


Figure 2.1: *Engagement histogram example.*

forms. While some exhibit bursty behaviour over longer time intervals, and others show patterns possibly arising from circadian rhythm, most conversations accumulate a large fraction of their total engagement only a few hours after posting. By visual inspection of the conversation histograms, it is clear that engagement rarely stays at a constant non-zero level at the resolution of 1 hour. Instead, peaks and fluctuations are common. To better understand the characteristics of the Twitter conversation we divide the temporal interaction patterns into four categories:

(1) conversations that have a single peak in interactions,

(2) two distinct peaks,

(3) one larger peak and one smaller peak,

(4) more than two peaks.

We will refer to these categories as (1), (2), (3), and (4), respectively. The single peak category is intended to group together conversations where there is a sharp rise and decay in engagement. This pattern is readily observed in the data - a post sparks some initial interest which quickly fades away. The remaining categories account for different ways in which engagement rises again after decaying. Conversations that have two prominent interaction peaks are grouped into the second category. When the second peak is small relative to the first peak, it could be reasonable to put it into the first category. For instance, the dip may be due to the circadian rhythm of humans using social media, creating an artificial peak in engagement that would not have occurred should the time of day have been different. The third category therefore groups conversations that have a smaller second peak, a grey zone between category (1) and (2). Finally, the fourth and most heterogeneous category collects the conversations that have more than two peaks, no matter the size. We admit that distinguishing what is a peak is in some sense subjective, nevertheless, we will describe how to define a peak, classify conversations into this taxonomy, and give more precise definitions of the categories in what follows.

To estimate the frequency of each category and the distribution of delay in engagement peaks, we use a method based on the algorithm by [16]. The algorithm is based on the idea of a significance test, and classifies values in a sequence as either a peak or non-peak based on whether the value exceeds a threshold computed from the mean and standard deviation of previous values in the sequence. The algorithm is described in greater detail in Section A.2. We use the result of the peak detection algorithm along with a few heuristics to discard peaks that are too small or too close to each other. We call this process Algorithm 2, and describe it in Table A.2.

After running Algorithm 2, a conversation is assigned to category (1) if there is only one peak, (2) if there are two peaks, and the smaller peak has a magnitude larger than 40% of the other, (3) if there are two peaks, and the smaller peak has a magnitude smaller than 40% of the other, and (4) otherwise. If the histogram has a peak at hour k to $k+1$, we refer to this interval as a *peak bin*.

2.5 Modelling Conversations

This section presents the models that will be used to describe interactions in Twitter conversations in time. We explain how the data from Twitter is transformed into a sequence of values to fit the model, as well as the model fitting and evaluation procedures.

2.5.1 Linear Models

A linear time-invariant system of the first order can be described by the relation

$$\dot{x} = \lambda x + \beta u \tag{2.1}$$

where $x = x(t) \in \mathbb{R}$ is the state vector, and $u = u(t)$ is the input signal. We define the Dirac delta $\delta(t)$ to satisfy

$$\int_{-\infty}^{\infty} \delta(t)dt = 1, \text{ and } \delta(t) = \begin{cases} +\infty & t = 0 \\ 0 & t \neq 0 \end{cases},$$

and the Heaviside function

$$H(t) = \int_{-\infty}^{t} \delta(\tau)d\tau = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0 \end{cases}.$$

Given an input signal $u = \delta(t)$, and assuming $x(t) = 0$ for $t \leq 0$, the system in 2.1 has the solution

$$x(t) = \beta e^{\lambda t} H(t). \tag{2.2}$$

We take the output to be $y(t) = x(t)$. Equation (2.1) describes the decay (assuming $\lambda < 0$) of x at a rate that is proportional to its magnitude. The time it takes for the quantity $y(t)$ to halve at time t is found by setting

$$y(t) = 2 \cdot y(t + T_{1/2}),$$

by which we obtain $T_{1/2} = -\frac{\ln 2}{\lambda}$. This holds for any $t > 0$, and $T_{1/2}$ is called half-life period, or half-life. Equation (2.1) is a deterministic model that can be used to approximate the decay of large discrete quantities, but not individual units. For instance, the decay of atoms is stochastic in nature [6], and thus, a more proper definition of half-life is the *expected* time it takes for a quantity to decay to halve.

The second order system

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ \rho \end{pmatrix} u(t), \tag{2.3}$$

with output $y = x_1 + x_2$ has been used in [4] to describe the decay of collective attention for a range of cultural items such as movies, music, scientific articles and more. The decay process is modelled by two mechanisms – communicative memory and cultural memory – with different decay rates, as the former feeds the latter. The idea is that an item initially receives attention by occupying a share of people's everyday conversations, eventually leading to the creation of cultural records (e.g., books or articles) about the item. The word-of-mouth diffusion eventually fades, as does the rate at which cultural records are produced. While the collective attention towards an item may remain constant, the relative share of attention can still shrink as total memory increases [4].

For $u(t) = \delta(t)$, the system (2.3) has the solution

$$x_1(t) = \frac{\rho\beta}{\alpha - \gamma}(e^{\alpha t} - e^{\gamma t})H(t) \tag{2.4}$$

$$x_2(t) = \rho e^{\gamma t} H(t) \tag{2.5}$$

and thus the output,

$$y(t) = x_1(t) + x_2(t) = \frac{\rho\beta}{\alpha - \gamma} e^{\alpha t} H(t) + \rho\Big(1 - \frac{\beta}{\alpha - \gamma}\Big) e^{\gamma t} H(t), \tag{2.6}$$

is a sum of two exponential functions. The detailed solution is found in Section A.3.

2.5.2 Model Fitting

We fit models of type (2.1) and (2.3) to conversation data that has been prepared in the following way. Using the histogram binning described in the previous section, for each conversation ζ_i we let the total number of interactions in the half-open interval $[k\Delta t, (k+1)\Delta t)$ be $y_i^D[k]$, $k = 0, ..., k_\mathrm{f}^i$, where $k_\mathrm{f}^i \in \mathbb{N}$ such that $(k_\mathrm{f}^i + 1)\Delta t$ is the smallest multiple of Δt that exceeds the final interaction time observed for ζ_i. As stated above, we use $\Delta t = 1$. We let the engagement described by a model during $[k\Delta t, (k+1)\Delta t)$ be $y_i[k] = y_i(k\Delta t)$. From (2.2) and (2.6), we have $y(0) = 0$, but in practice we let $y_i[0] = y_i(0^+) > 0$.

When fitting a model we append K bins with value zero to the engagement histogram, i.e., $y_i^D[k] = 0$ for $k = k_\mathrm{f}^i + 1, \ldots, k_\mathrm{f}^i + K$, to avoid estimating an unrealistically slow decay. This can typically happen when k_f^i is small. For instance, if $y_i^D[k]$ is non-zero for k close to k_f^i, we give a penalty for letting $y_i[k]$ be non-zero also at times $k = k_\mathrm{f}^i + 1, \ldots, k_\mathrm{f}^i + K$ when there are no interactions. We set K so that $k_\mathrm{f}^i + K \geq 72$.

As model fitting criterion we use the mean square error (MSE) loss between the model and the data

$$\mathrm{MSE}(y_i, y_i^D) = \frac{1}{k_\mathrm{f}^i + K + 1} \sum_{k=0}^{K+k_\mathrm{f}^i} \varepsilon_i^2(k;\theta), \tag{2.7}$$

where $\varepsilon_i(k;\theta) = y_i[k] - y_i^D[k]$ is the error for the interval $[k, k+1)$, and θ represents the model parameters.

2.5.3 Model Evaluation

In order to assess which of the two models introduced in Section 2.5.1 is more suitable for describing Twitter interactions we use the Akaike information criterion (AIC) [10]:

$$\mathrm{AIC} = \min_\theta \left(1 + \frac{2\kappa}{N}\right) \sum_{k=0}^{N} \varepsilon_i^2(k;\theta), \tag{2.8}$$

with $N = K + k_\mathrm{f}^i$. For the first order system we have $\kappa = 2$ parameters, and for the second order system, $\kappa = 4$.

The fit of the model is evaluated using the normalised residual sum of squares (RSS):

$$R_i = \frac{\sum_{k=0}^{N} \varepsilon_i^2(k;\theta)}{\sum_{k=0}^{N} (y_i^D[k])^2}. \tag{2.9}$$

3 Results

This chapter contains the results of our analysis. First, we give an overview of the dataset along with some basic statistics on the collected conversations. We then present the results of fitting the models to engagement data, and compare them to see which is more suitable for describing Twitter dynamics. Finally, we investigate what impact the interactions of users with a high follower count might have on peaks in the conversations.

3.1 Conversation Statistics

Processing the conversations with Algorithm 2 and following the taxonomy in Section 2.4.1, the conversations are divided into categories according to Table 3.1, additionally separating the cases with a peak during the first hour (no delay), or if it arrives later (delay). Using this categorisation, we find that each conversation belongs to exactly one group. Since there is no ground truth, the validity of the results is verified to be reasonable by visual inspection. Several observations can be made from the categorisation result. We note that close to three quarters of all conversations have a dominant single peak, counting both category (1) and (3), described in Section 2.4.1, as obeying this pattern.

In about a third of conversations there is a delay between the root tweet post and the arrival of interactions, or at least a slower accumulation of interactions. If the delay is too large, it cannot adequately be modelled by the first order system (2.1).

Moreover, we find that the tweet volume proportions between the categories do not differ notably from the class proportions as seen in Figure 3.1. The largest difference is that conversations with more than two peaks make up a larger share of the total tweet volume, while the tweet volume of single peak conversations shrinks in proportion. Figure 3.3 shows that the tweet volume distributions are

Table 3.1: Number of conversations in each category as described in Section 2.4.1.

	No delay	Delay	Total
Single peak	4599	1521	6120 (62.7%)
Double peak	532	448	980 (10.0%)
Double peak (small 2nd)	813	400	1213 (12.4%)
Multiple peaks	943	502	1445 (14.8%)
Total	6887 (70.6%)	2871 (29.4%)	9758 (100.0%)

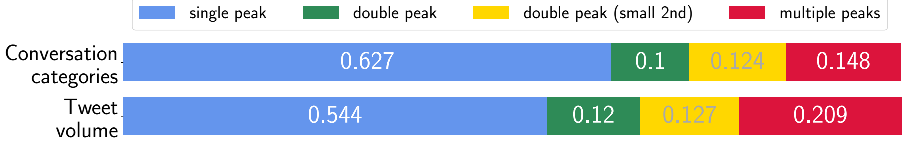


Figure 3.1: Proportion of tweet volume and conversations for the different categories described in Section 2.4.1.

also relatively stable after the categorisation. The tweet volume distribution for multiple peak conversations has a somewhat heavier tail. The proportion of tweet types in the whole dataset is shown in Figure 3.2. It can be noted that retweets, the interaction type that requires the least effort, make up the majority (80%) of tweet volume.

3.1.1 Circadian Rhythm

It is not surprising that the living patterns of humans affects how content is posted on social media platforms. In addition to fluctuations due to daily rhythm, seasonal and geographical patterns can be found as well [9]. As previously noted, there are periodic patterns in our dataset, posing the question whether the secondary peaks or the delays in engagement can be linked to periodicity in human activity at the conversation level.

We sample tweets from all of Twitter, meaning that we obtain data from users in many different time zones. Tweet payloads contain only the time stamp in UTC

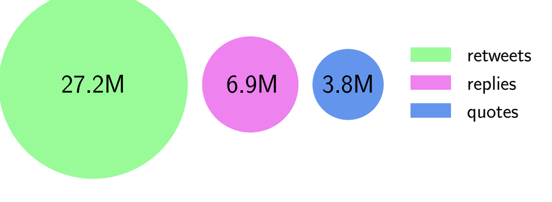
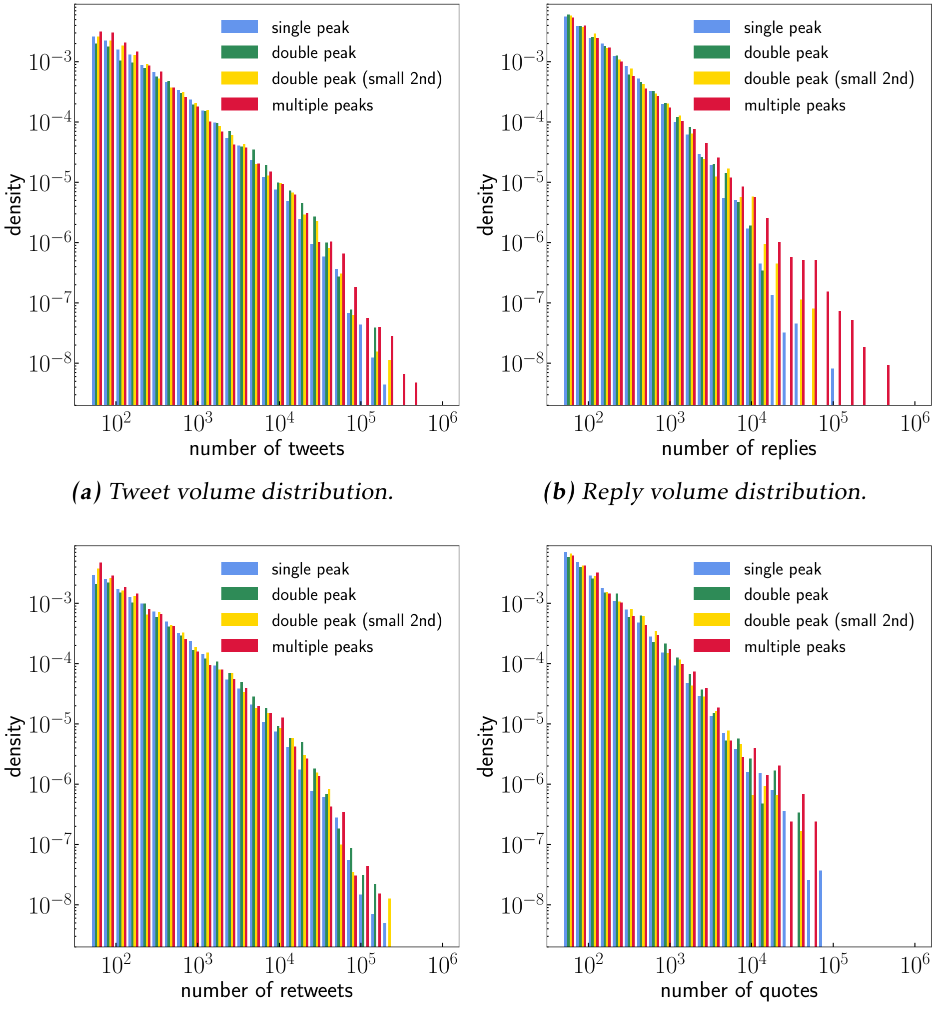


Figure 3.2: Number of tweets of each type in the dataset.



***(a)** Tweet volume distribution.*

***(b)** Reply volume distribution.*

***(c)** Retweet volume distribution.*

***(d)** Quote tweets volume distribution.*

***Figure 3.3:** Tweet volume distributions for each conversation category.*

time, geolocation data is sparse, and languages such as English and Spanish are used in several different time zones. With this limitation, modelling the effect of circadian rhythm on interactions is beyond the scope of this thesis, but for illustration purposes we try to visualize the effect by looking at a subset of our data.

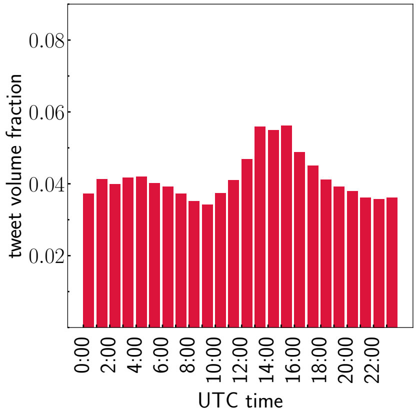
We filter root tweets by language, extracting posts in English, Spanish and Japanese from the dataset. For each language set we sample tweets so that their posting time is evenly spread out over the 24 hours of a day. We then take the replies, retweets and quotes of these roots, forming one set of tweets for each language. This amounts to 13,014,847 English tweets from 4,745 conversations, 1,650,516 Spanish tweets from 1,301 conversations, and 620,567 Japanese tweets from 469 conversations. The aggregation of the posting time is shown in Figure 3.4.

For tweets in English, we see a more even distribution compared to Spanish and Japanese. There is a small increase in posts at UTC 03:00, corresponding to evening of U.S. time zones, and a larger increase at UTC 14:00, evening time in India. Spanish tweet volumes show a dip corresponding to the night-time of UTC-5, the time zone of U.S. east coast and Latin America. Japanese tweets clearly exhibit patterns of a daily rhythm: one peak during noon, and one during the evening, and a visible dip during night-time.

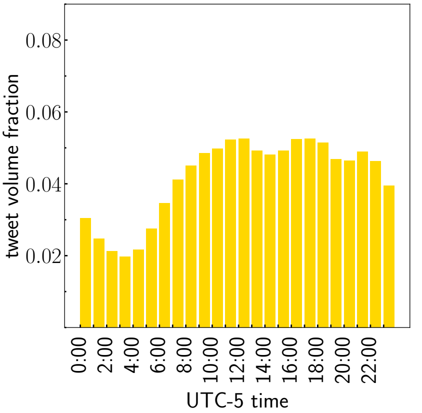
3.2 Temporal Model

The AIC of both first and second order models introduced in Section 2.5.1 is evaluated for each conversation, and the result is shown in Table 3.2. We find that in over 95% of cases the second order model is superior in terms of AIC. Table 3.3 shows the normalised RSS for all conversation types, with an additional split on whether there is a peak during the first hour (no delay), or if it arrives later (delay). Figures A.1-A.8 in Section A.5 show the fit of the models for different conversation types. Being sums of exponential functions, the models (2.1) and (2.3) are better suited to describe the exponential decay of engagement often found in categories (1) and (3), which is confirmed by looking at the normalised RSS for these conversations, but only for the case of no delay. Both models show a poor fit to the conversations where the initial peak is delayed compared to when there is no delay.

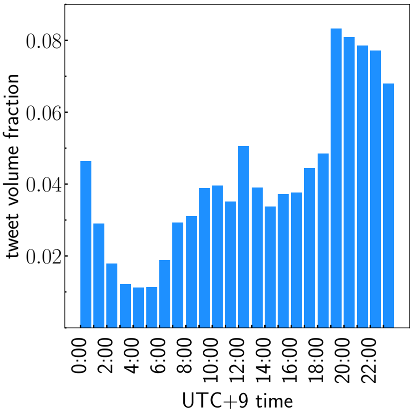
We investigate the time constants $\tau_\alpha = -1/\alpha$ and $\tau_\gamma = -1/\gamma$ for the conversations of type (1)-(3) with no delay in the initial peak, to which the fitted model has a normalised RSS lower than 0.2. Denote this set of conversations by Γ. This corresponds to 57% of fitted conversations. The distribution of the time constants of conversations in Γ is shown in Figure 3.5 and the percentiles of the distributions are presented in Table 3.5. Bins spanning 1 hour are used for fitting the model, however, because the rapid decay in engagement frequently plays out in mere minutes, we also estimate τ_γ using 5 and 15 minute bins, while keeping the 1 hour estimate as a reference.



(a) English tweets. Time axis in UTC0.



(b) Spanish tweets. Time axis in UTC-5 (U.S., Latin and Central America time zones).



(c) Japanese tweets. Time axis in UTC+9 (Japan Standard Time).

Figure 3.4: *Tweet volumes of different languages during the day.*

Table 3.2: *Fraction of conversations where the AIC* (2.8) *of the second order system* (2.3) *is lower than that of the first order system* (2.1).

Category	Fraction
Single peak	0.967
Double peak	0.904
Double peak (small 2nd)	0.957
Multiple peaks	0.923
All	0.951

Table 3.3: *Mean and median normalised residual sum of squares of both models for each conversation type as defined in Section 2.4.1.*

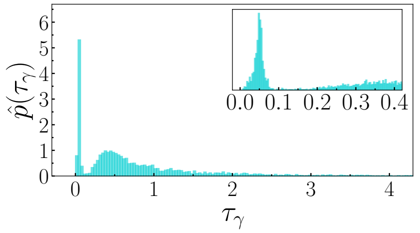
Category	**First order model**		**Second order model**	
	mean	median	mean	median
Single peak, no delay	0.075	0.050	0.035	0.020
Double peak, no delay	0.405	0.353	0.293	0.271
Double peak (small 2nd), no delay	0.268	0.181	0.170	0.106
Multiple peaks, no delay	0.480	0.433	0.356	0.318
Single peak, delay	0.374	0.291	0.254	0.174
Double peak, delay	0.413	0.373	0.296	0.268
Double peak (small 2nd), delay	0.419	0.387	0.310	0.265
Multiple peaks, delay	0.505	0.463	0.413	0.371

Table 3.4: *Normalised residual sum of squares of both models for different bin widths.*

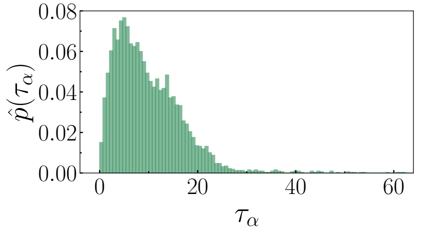
Bin width	**First order model**		**Second order model**	
	mean	median	mean	median
5 min	0.439	0.392	0.348	0.302
15 min	0.341	0.265	0.251	0.177
1h	0.252	0.155	0.175	0.089
2h	0.214	0.112	0.144	0.059

Table 3.5: *Percentiles of the distributions of the estimated time constants, τ_γ, and τ_α, for second order systems fitted to conversations of category (1)-(3) (no delay). The second and third rows show percentiles for the distributions of τ_γ computed using 15, and 5 minute bins respectively.*

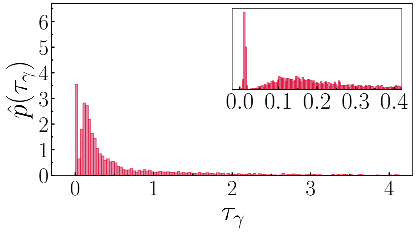
	Percentile				
	5:th	25:th	50:th	75:th	95:th
τ_γ	0.041	0.263	0.551	1.067	3.085
τ_γ (15 min)	0.012	0.117	0.225	0.565	2.359
τ_γ (5 min)	0.024	0.062	0.121	0.423	2.269
τ_α	1.658	4.689	8.451	13.753	21.791



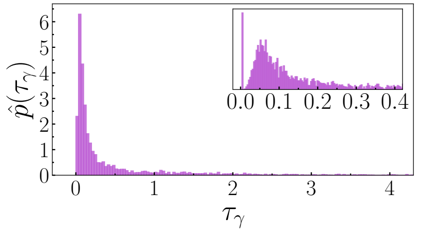
(a)



(b)



(c)



(d)

Figure 3.5: *Estimated distributions of the time constants, (a) τ_γ and, (b) τ_α, for second order systems fitted to conversations of category (1)-(3) (no delay) using 1 hour bins. Below, we show the distributions for τ_γ for the same category of conversations estimated with (c) 15-minute bins and (d) 5-minute bins. The inset figures in (a), (c) and (d) are zoomed-in versions of the respective histograms.*

3.3 Influencer Analysis

A natural assumption to make is that a user with a large number of followers, hereafter referred to as an *influencer*, has potential to spread a tweet to many more users. In general, however, it seems that the number of followers is a poor measure of a user's potential of spreading tweets and ideas to other users. The findings in [5] suggest that having many followers alone is no guarantee for generating many interactions towards one's content. In [15], the follower count of the root author was not found to be a reliable predictor of total engagement (retweets) towards a tweet.

In this study, we are asking whether influencers affect the conversation dynamics. In other words, is it likely that an interaction from an influencer will trigger an increase in engagement?

There are multiple ways to define an influencer. In particular, it is useful to consider two definitions. First, for each conversation, we define *local* influencers as the users with a large following relative to other users in the same conversation. Second, we define *global* influencers as users that have a follower count which can be considered large in the overall compendium of users in our dataset.

We let the threshold for including a user in the local influencer group vary by taking the top 1, 5, and 20 users, and users with a follower count in the 99:th, 95:th, 90:th, and 80:th percentiles for that conversation, giving in total 7 sets of influencers. The mean and median number of detected influencers over all conversations per threshold choice are shown in Table 3.6. When selecting the top 1, 5, and 20 influencers, we avoid choosing users with low follower count by applying the 95:th percentile as a lower bound on our threshold. Hence, the number of influencers in the sets of top 1, 5, and 20 influencers may be less than 1, 5 and 20 respectively.

We begin our analysis by noting that local influencers do appear in conjunction with peaks. Taking the top 5 influencers, 58% of all detected peaks have an influencer on the peak bin or on bin prior to the peak bin, compared to 32% in the same period after the peak bin. For the conversation categories, the proportions are (1): 92%, (2): 64%, (3): 62%, and (4): 31%. The frequency increases when excluding peaks that are small in comparison to the largest peak in the conversation. If we include only peaks of magnitude 0.7 times the maximum peak height, influencers are then present in 81% of cases. One should note that peak height depends on the choice of bin positions and is not a perfect measure of the engagement.

To see whether an influencer interaction generates above-average engagement, we compare the engagement accumulated by the conversation in a window before and after the interaction. Engagement in the window after an influencer interaction is denoted E_a, and engagement in the window before is denoted E_b. We consider the interactions of replying, retweeting, and quoting separately, since the impact on the followers' feeds is different. In the following analysis, the root is not counted as a reply in the case that the author is considered an influencer.

We look at the fraction of influencer interactions for which $E_b < E_a$ holds, measuring in a window around the influencer interaction. For a window size of

Table 3.6: Mean and median number of influencers detected in each conversation by thresholding choice.

Threshold choice	mean	median
80:th percentile	652.4	111
90:th percentile	326.7	56
95:th percentile	163.7	28
99:th percentile	33.4	6
top 20	15.1	20
top 5	4.8	5
top 1	1.0	1

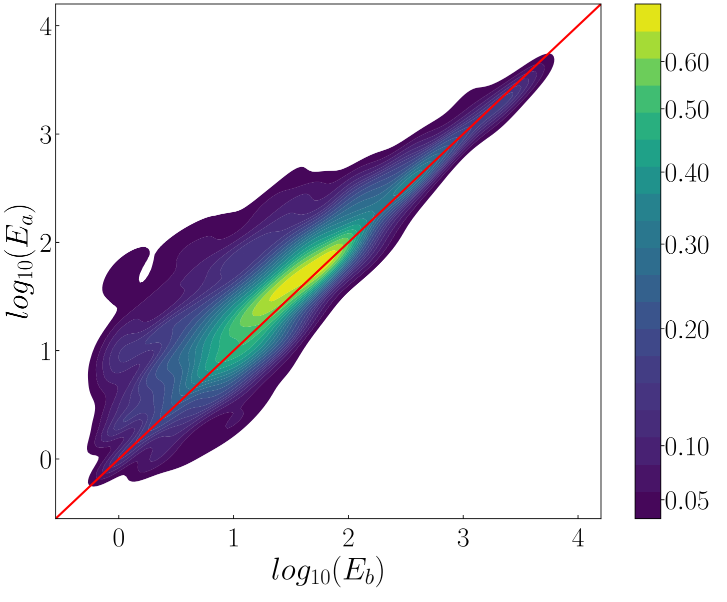


Figure 3.6: *Scatter plot of* $\log_{10}(E_a)$ *and* $\log_{10}(E_b)$ *using a kernel density estimate for quotes from influencers in the top 1 set of local influencers, computed using a window size of 15 minutes. The red line shows* $E_a = E_b$*.*

15 minutes, taking influencers from the top 1-set as defined above, the fraction of influencer quotes for which $E_b < E_a$ is 0.67, while for $E_b > E_a$ it is 0.27. This means that in 67% of influencer quote appearances, the conversation engagement in the 15-minute window following the quote was higher than in the same period prior to the quote. The opposite ($E_b > E_a$) is true in 27% of cases, and $E_b = E_a$ otherwise. Figure 3.6 shows a scatter plot of E_a and E_b using a kernel density estimate on log-transformed engagement computed using a window size of 15 minutes. We see that in the range $E_a, E_b < 100$, the density is higher above the line $E_a = E_b$. Figures 3.7a-3.7c show how the proportions of cases for which $E_b < E_a$ change when varying the size of the window and the set of influencers for replies, retweets and quotes respectively. The fraction of cases where $E_b < E_a$ for an influencer quote or retweet shrinks when increasing the window size, indicating that the effect diminishes quite rapidly. Including more users in the influencer set also decreases the fraction. There is no clear indication that influencers replies affect engagement in the same way as retweets and quotes.

To quantify this further, we compute the Spearman rank-order correlation coefficient between the difference in number of interactions before and after an influencer reply, retweet and quote, and the number of followers of the same influencer. That is, we compute the rank correlation between $E_a - E_b$ and the followers of the corresponding influencers. In Figures 3.8a-3.8c we see a similar pattern as in Figures 3.7a-3.7c. For retweets and quotes, the positive correlation between $E_a - E_b$ and the follower count is present only when we are restrictive with which users are included in the influencer set and using a smaller window, and for replies we cannot find a relation.

Choosing instead to define which users are influencers on a global scale, we use thresholds corresponding to the q:th percentile for $q \in \{90, 95, 99, 99.5, 99.9, 99.99\}$ of the follower counts of all 33.9M users over all sampled conversations (including the users appearing in conversations with fewer than 50 interactions). This is not a random sample of the follower distribution on Twitter as the users that seldom interact have a lower probability of being sampled. Nonetheless, the obtained distribution is heavy-tailed, appears to obey a power law with exponent 2.1, and exhibits the signs of low-degree saturation and high degree cut-off that is common to real world networks [3, Ch. 4]. In Figure 3.9, we plot the Spearman correlation between influencer followers and $E_a - E_b$, and the fractions of influencer interactions for which $E_b < E_a$, for the set of influencers defined using the global thresholds and for varying window sizes. We find only a weak positive correlation for quote interactions in this case (Figure 3.9f), although for most quotes and retweets of influencers we have $E_b < E_a$ (Figures 3.9a and 3.9c).

With the present dataset we cannot raise the threshold in the global definition of an influencer if we want to retain statistical significance. For reference, setting $q > 99.995$ corresponds roughly to picking the top 1-5 users with most followers in each conversation, with the additional condition that they have at least 1,000,000, followers, which leaves a few hundred data points to divide between replies, retweets, and quotes.

With the results from using the global influencer definition in mind, we return to investigate the findings using the local definition of an influencer. Since

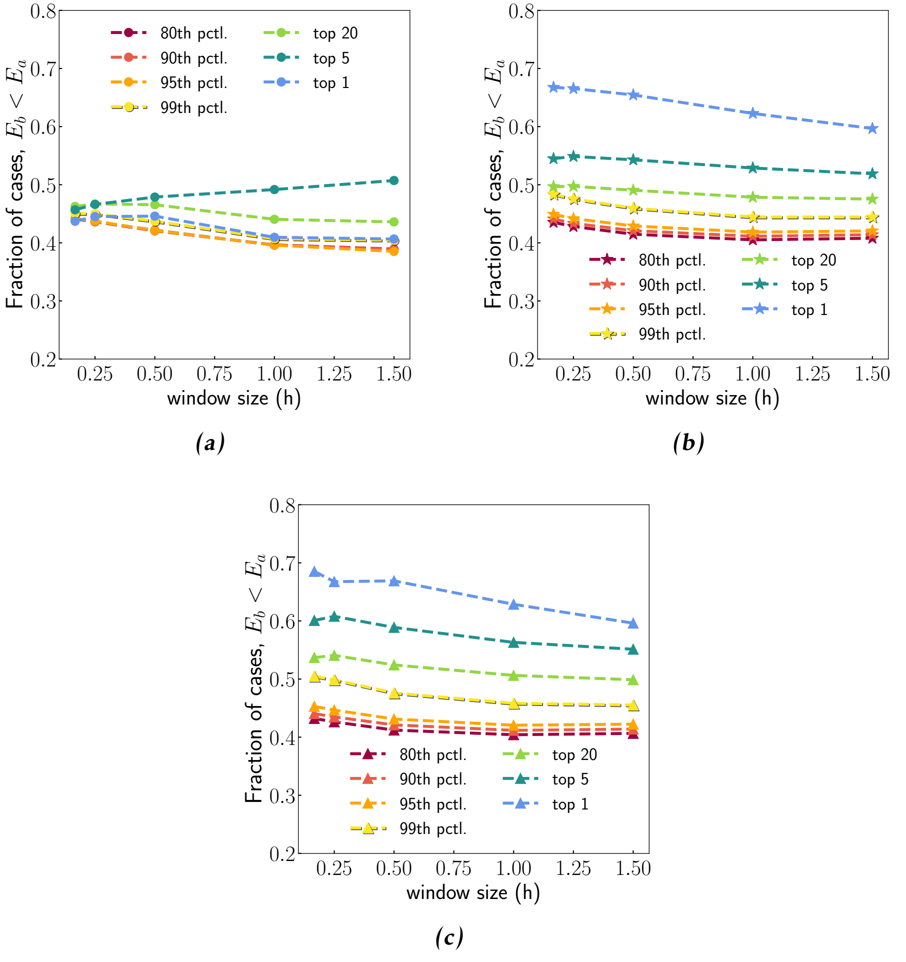


Figure 3.7: *Fraction of influencer interactions with more engagement afterwards than before for varying window sizes and influencer sets, where influencer interaction is (a) replies, (b) retweets, (a) quotes.*

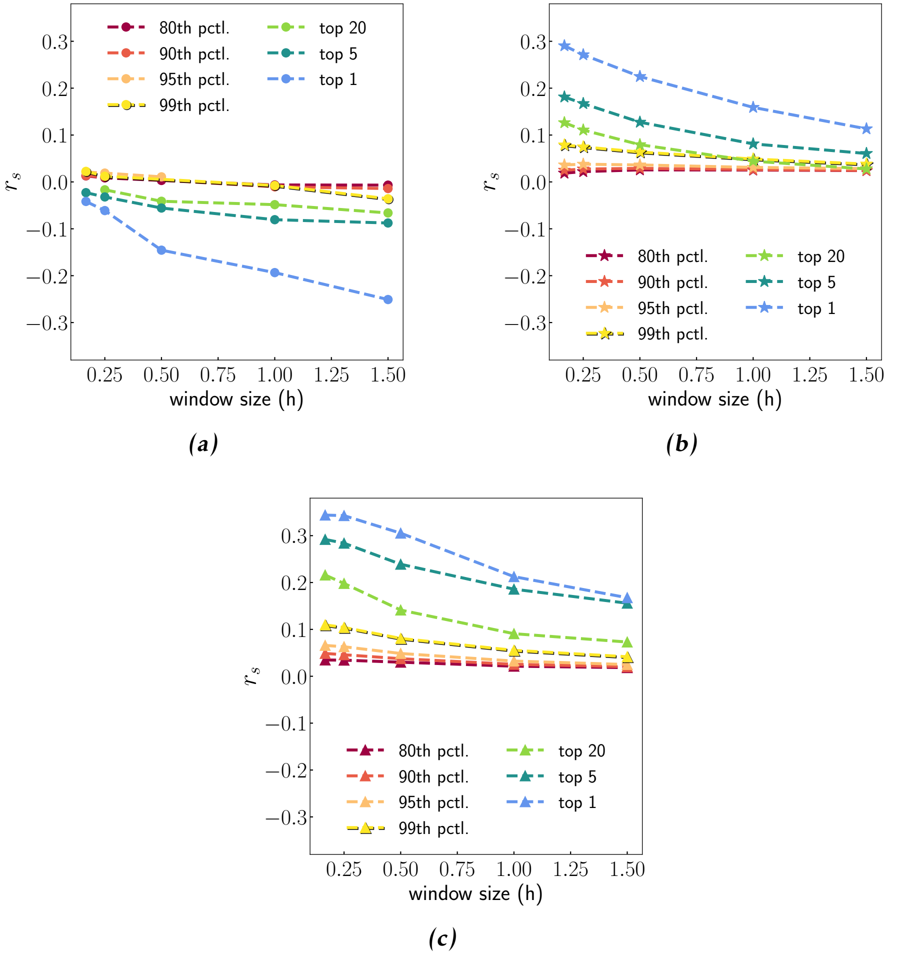


Figure 3.8: *Spearman rank correlation, r_s, between $E_a - E_b$ and influencer follower count for varying window sizes and sets of local influencers, where the influencer interaction is a (a) reply, (b) retweet, (c) quote.*

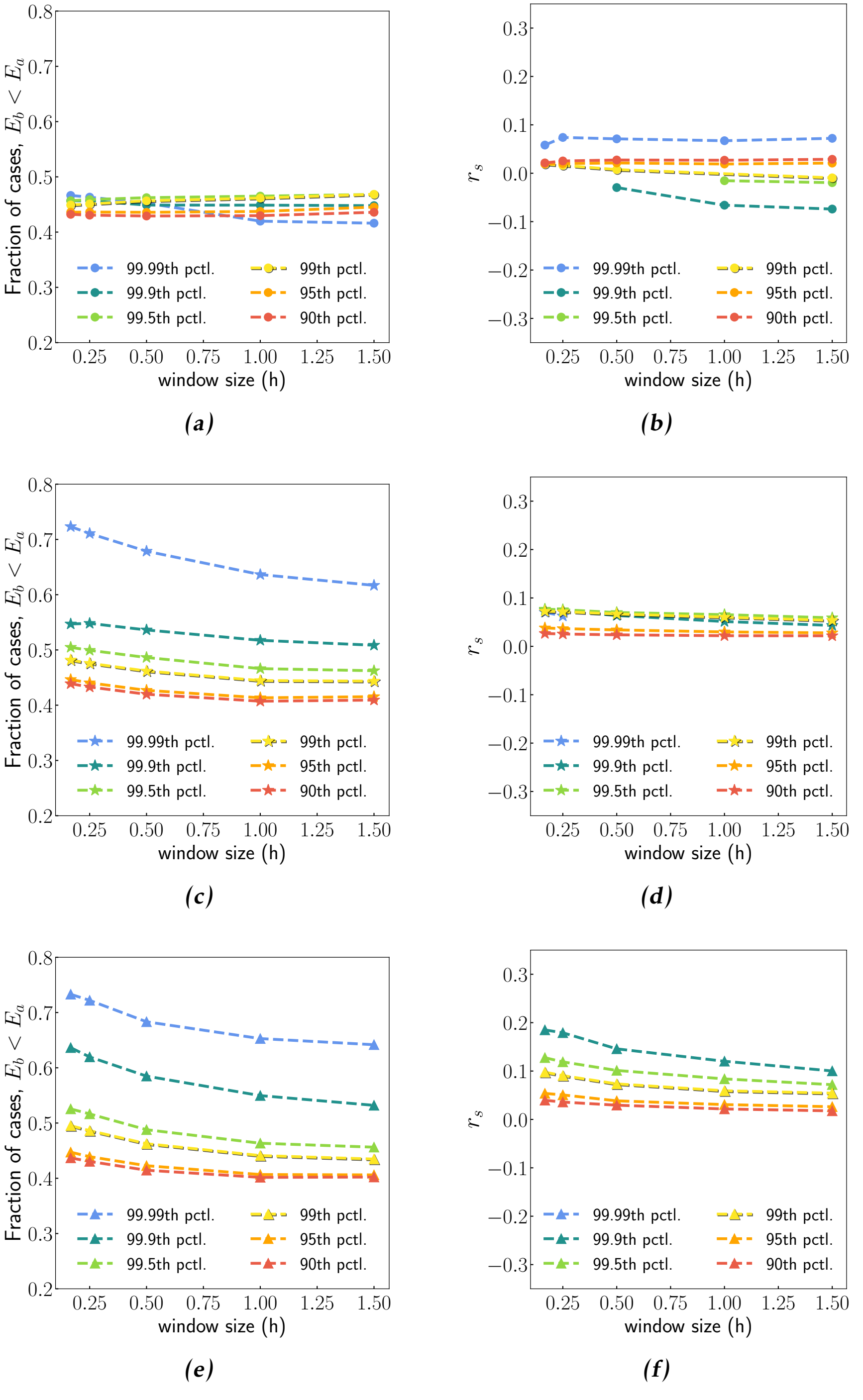


Figure 3.9: *(a), (c), (e): Fraction of global influencer interactions for which $E_a > E_b$, where influencer interaction are replies, retweets, and quotes, respectively. (b), (d), (f): Spearman rank correlation, r_s, between $E_a - E_b$ and influencer follower count for varying window sizes and sets of global influencers, where the influencer interactions are replies, retweets, and quotes, respectively. Statistically insignificant data points have been excluded.*

Table 3.7: *Percentiles of the global follower count distribution of our dataset, rounded to the nearest thousand.*

Percentile	threshold
90	1,300
95	2,500
99	10,000
99.5	18,000
99.9	89,000
99.99	630,000
99.995	1,000,000
99.999	5,389,000

we have a weak positive correlation between number of followers and $E_a - E_b$ for influencers with a follower count below the 99:th percentile globally, we investigate whether the correlation is present for local influencers with few followers in the global sense. We find that the effect disappears when considering local influencers with fewer than 20,000 followers (corresponding to users below the 99:th percentile of the global follower count distribution). Instead, we combine the local and global definitions, imposing the condition that an influencer should have a follower count that is larger than 20,000 and is in the right tail of the distribution for the conversation. For quotes and retweets, this results in a Spearman correlation higher than what we had when employing only the global definition. Plots of the correlation between $E_a - E_b$ and influencer followers for varying window sizes are shown in Figure 3.10.

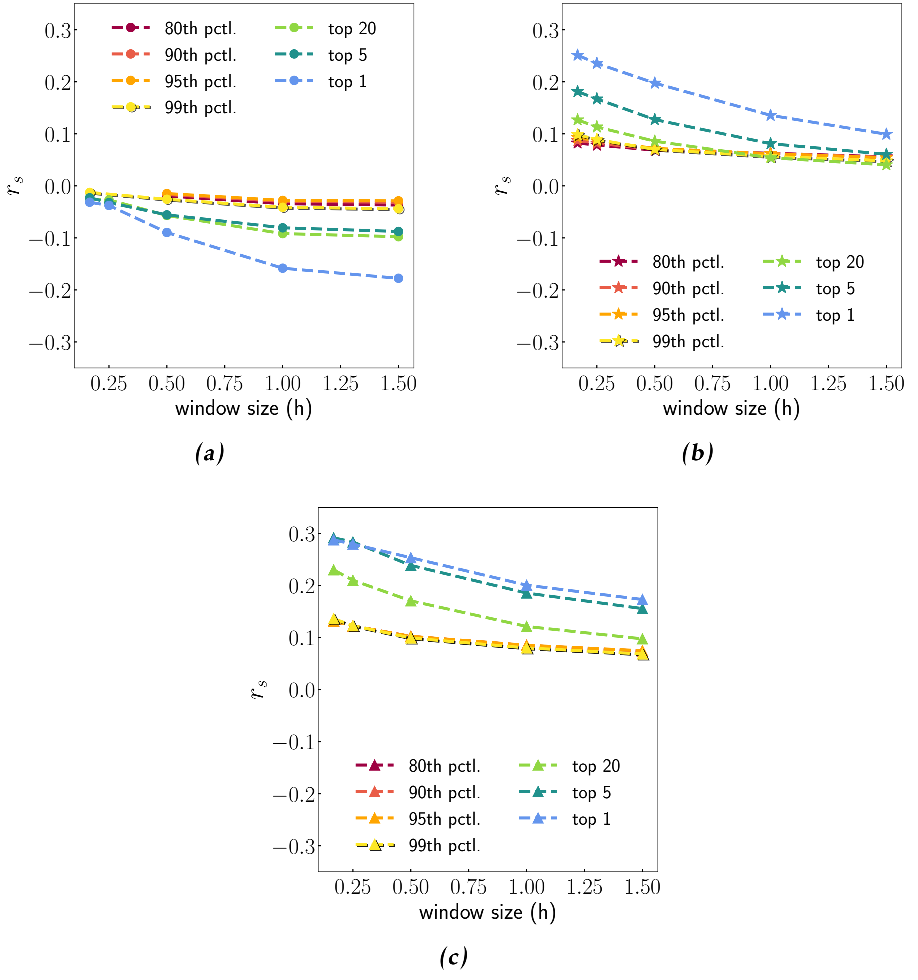


Figure 3.10: *Spearman rank correlation, r_s, between $E_a - E_b$ and influencer follower count for varying window sizes and sets of influencers (local and global), where the influencer interaction is a (a) reply, (b) retweet, (c) quote.*

4

Discussion

4.1 Temporal Models and Conversation Time Constants

We see in Section 3.2 that the second order model is preferable in terms of AIC when modelling conversation engagement. The superiority of the second order model suggests that conversation dynamics is better modelled by two time constants, which describe the two phases of the conversation: a sharp peak with a rapid initial decay in engagement followed by a period of low intensity in interactions. This can be seen in Figures A.1 and A.3. The descriptive power of the second order system, not to mention the first order system, decreases as multiple peaks and large delays are introduced. This is no surprise, as the implicit assumption that engagement arrives directly after a post, at least in small quantities, does not hold, which can be seen in the top right time series in Figure A.7. Delaying the input impulse to the system improves the performance of both models, but requires another variable or process, such as a peak detection algorithm, and does not resolve the problem of multiple peaks.

Recall that we used Γ to denote the set of conversations of type (1)-(3) with no delay in the initial peak. In order to understand what the results in Table 3.5 imply for our model, let us assume that the output of the second order system (2.6) is monotonically decreasing (for all $t > 0$ when $u(t) = \delta(t)$), as should be the case for conversations in Γ. Whenever τ_γ is small relative to τ_α, the initial decay of engagement is comparatively fast, and the half-life[1] of the conversation in the late phase (i.e., $e^{\gamma t} \approx 0$) will be $T^{\alpha}_{1/2} = \tau_\alpha \ln 2$. The half-life should in this context be understood as the period within which the intensity in engagement is halved. The dynamics of the initial phase (when $e^{\alpha t} \approx const.$) is to a larger extent

[1] strictly speaking, the term half-life is not applied to second order systems, but we consider $e^{\gamma t}$ to be very small in the late phase

governed by τ_γ. However, the decay rate also depends on the weight that governs the proportion of the exponential terms: $\eta = \frac{\beta}{\alpha-\gamma}$. As can be seen in (2.6), when η is close to zero, the term containing $e^{\gamma t}$ dominates, and when it is close to unity, $e^{\alpha t}$ dominates. In the latter case, we get a solution that behaves similar to a single exponential, with an actual half-life $\tau_\alpha \ln 2$ for the entire span of the conversation. In the limit as τ_γ approaches τ_α, the second order model has a response identical to a single exponential term.

For conversations with a delayed initial peak, it is feasible that a fitted second order model will have $\eta > 1$, and the response is not monotonically decreasing. Instead it increases from $t = 0$ and reaches a peak at time $t = \frac{1}{\alpha-\gamma} \ln \frac{\beta\alpha}{\gamma(\beta+\gamma-\alpha)}$, after which it decreases. A small time constant τ_γ in this setting implies a rapid rise.

Looking closer at Table 3.5, we see that the median of τ_γ for conversations in Γ is 0.225 and 0.121 for estimates using 15 and 5 minute bins, respectively. This implies that the term with $e^{\gamma t}$ in (2.6) halves every 0.156 and 0.084 hours (9.3 and 5 minutes) in the respective case. These periods are very close to the bin size, which begs the question how reliable the estimates are. By narrowing the bin size further we risk ending up with more bins containing no interactions, which may favour a solution that decays more quickly. In Table 3.4, we also see a larger normalised RSS for the model fits using 5-, and 15-minute bins compared to 1 hour bins. The estimates of τ_α are likely more accurate since the time scale is larger than the bin sizes.

The typical late phase half-life, $T^{\alpha}_{1/2}$, for conversations in Γ is around 3-10 hours, which indicates that conversations still accumulate engagement even after some hours, although the rate is in general lower. For the initial phase we see fast dynamics - the time constant τ_γ typically being in the order of minutes rather than hours - showing that the pace on Twitter is very fast indeed.

4.2 Impact of Influencers

In Section 3.3, we investigate the impact of influencers on conversation dynamics. The results suggest that an influencer sharing a tweet, either by a quote or retweet, is linked to a rise in conversation engagement. The first observation we make is that influencers more commonly appear before peaks rather than after, and that the number of interactions more often than not are higher after an influencer retweet or quote, than before. We also find a positive correlation between $E_a - E_b$ and the number of followers of the influencer that retweeted or quoted the conversation root. The results hold when defining an influencer to be a user that has many followers compared to the set of users in the same conversation, as well as compared to all other users in our dataset. Moreover, these results are robust to change in the size of the window in which we count user interactions.

Naturally, influencers are not the only driving force behind the rise and decline of engagement rates. As the authors of [5] and [15] point out, having many followers is not a solid guarantee for spawning interactions. Defining an influencer as having many followers in a global context, a high threshold (>600,000) was needed to obtain results similar to combined local and global definitions. It makes sense that applying the local and global definition of an influencer in combination enhances the effect on interactions, since a retweet of an influencer with 10,000 followers may have a relatively small impact on interactions if an influencer with 100,000 followers also retweets the conversation root. Apart from influencers, there are many other possible factors that affect interaction patterns. An external event can suddenly make a conversation topic interesting, or it could be due to the daily rhythm of users on Twitter, as explored in Section 3.1.1. We are also agnostic of how the Twitter algorithm decides what content to present to users.

In summary, we find it likely that influencers' retweets and quotes contribute to the emergence of peaks in conversation interaction patterns. Knowledge of the user network could be used to further verify this effect. If the users that interact with the conversation after the influencer retweet or quote are followers, it is more likely that their presence is due to the influencer. Influencer replies seem to have no noticeable or consistent effect on engagement, which is consistent with our belief of how tweets become visible on the platform.

4.3 Future Work

There were many tracks during the course of this work that could not be pursued due to time constraints but are viable options for future investigations. Relating the topology of the Twitter network to conversation dynamics could give further insights into the pace and reach of information on the platform. As an example, one could examine whether tweets that propagate further away from the root author or into new clusters of users have an effect on the engagement the conversation receives. It is possible that the delay in interactions observed from the time of a post to the first peak in some conversations is related to the user network topology. As mentioned above, knowledge of the user network could also prove useful when evaluating the effect of influencer interactions on engagement. At present, the rate limit on collecting user follower lists is a considerable obstacle to any enquiry for which the user network is required – an alternative would be to use an older dataset.

In this study we aimed to collect a dataset that would be as general as possible with respect to types of user behaviours and interaction patterns. A direction to take from here would be to apply the analysis in this work to specific topics or user sets to investigate how communities differ in their use of Twitter, similar to the work of [17].

Appendix

A

Supplementary Information and Theory

A.1 Missing Data

As stated in Section 2.4, there is commonly a discrepancy in the number of tweets retrieved compared to the number in the `public_metrics` attribute of the tweet payload. For quotes and replies, this figure is small, and we often retrieve more tweets than the public metric attribute states are available. However, we sometimes retrieve fewer retweets than what is available. The problem mainly lies in the method of collecting retweets.

Of all tweet types, retweets is the least straightforward to retrieve. We rely on the full archive search endpoint for this task. Using the query commands `retweets_of:user_ID` and `"retweet text"` we match (re)tweets of tweets from the specified user ID that contain the phrase within the quotation marks. The phrase to match is picked from the root tweet text attribute, and is cleaned of special characters and truncated at the first hyperlink. The tweet text attribute of a root contains a shortened hyperlink to the tweet in question that is not visible when visiting Twitter. In general, the link is also part of the retweet text, but in some cases, it is missing. When this happens, sending a query on the entire tweet text returns no results from the full archive search endpoint.

Another issue of retrieving retweets is that the tweet text may match multiple tweets, yielding an avalanche of results that take a long time to retrieve, which fills up the tweet cap quickly. When mining data at scale, it is not uncommon to see users that post the same text every day or every hour. Examples of these types of accounts include those that advertise artists, fan pages, or users with regular giveaways. Comparing the number of queries and the total number of retweets gives an indication of when it is suitable to stop sending queries for the retweets of a certain tweet. Another not too rare outcome is to get no results at all, which may be due to a mismatch between the tweet text and the retweet text, or deletion

Table A.1: Peak detection algorithm proposed in [16].

Algorithm 1: Peak detection algorithm

Input: sequence $\{x_n\}_{n=0}^{N-1}$, and parameters ℓ, d, w

1: Compute

$$\mu_{\ell-1} = \frac{1}{\ell}\sum_{j=0}^{\ell-1} x_j, \text{ and } \sigma_{\ell-1} = \sqrt{\frac{1}{\ell}\sum_{j=0}^{\ell-1}(x_j - \mu_{\ell-1})^2},$$

and let $x_i^{\mathrm{f}} = x_i$, $i = 0, \ldots, \ell-1$.

2: **for** $i = \ell, \ldots, N-1$ **do**

  if $x_i \notin [\mu_{i-1} \pm d\sigma_{i-1}]$

    set $r_i = \pm 1$, and $x_i^{\mathrm{f}} = w x_i + (1-w)x_{i-1}^{f}$

  else

    set $r_i = 0$ and $x_i^{\mathrm{f}} = x_i$

  end if

$$\mu_i = \frac{1}{\ell}\sum_{j=i+1-\ell}^{i} x_j^{\mathrm{f}}$$

$$\sigma_i = \sqrt{\frac{1}{\ell}\sum_{j=i+1-\ell}^{i}(x_j^{\mathrm{f}} - \mu_i)^2}$$

end for

3: **return** $r = (r_\ell, \ldots, r_N)$

of the tweet or user.

A.2 Peak Detection Algorithm

We use a peak detection algorithm proposed by [16] to categorise Twitter conversations. Given a sequence of values the algorithm measures the deviation from the previous points using a Z-score test to determine if a value in the sequence is a peak. At each step, the algorithm considers the ℓ most recent data points and computes their mean μ and standard deviation σ. A peak is detected whenever the value of the next data point lies outside of the interval $\mu \pm d\sigma$, where d is a threshold parameter. In other words, a peak is a data point more than d standard deviations away from the mean or median of the ℓ most recent data points. The algorithm uses an influence parameter w to reduce the contribution of peaks to the mean. Setting w to zero will assume that the data with no signal present will have values between $\mu_{init} \pm d\sigma_{init}$, and any points outside that interval is considered a peak. The algorithm is formalised in Table A.1.

We make a few modifications of this algorithm which are summarised in Table A.2. First, we ignore any peaks detected when there is a dip in interactions ($r_i = -1$). The original algorithm would detect a sharp decay and subsequent rise of engagement as a peak in the opposite direction, which we are not interested in. Secondly, we ignore any detected peaks that have a magnitude lower than a frac-

Table A.2: *Extended peak detection algorithm for finding peaks in engagement histograms.*

Algorithm 2: Extended peak detection algorithm

Input: sequence $\{x_n\}_{n=0}^{N-1}$, and parameters ℓ, d, w, v, d_{adj}

1: Run Algorithm 1 to compute $r = (r_\ell, ..., r_N)$.

2: Let $r^{\text{mod}} = (r_\ell, ..., r_N)$.

3: Set $r_i^{\text{mod}} = 0$ for any i for which $x_i < v \cdot \max_{\ell \leq j \leq N}(x_j)$ or $r_i = -1$.

4: Find all maximal sets of peaks such that no peak occurs further away than d_{adj} steps. Remove all peaks in each set except for the one with the largest magnitude (set the corresponding element in r^{mod} to 0 for the peaks in question).

5: **return** $r^{\text{mod}} = (r_\ell^{\text{mod}}, ..., r_N^{\text{mod}})$.

tion, $v \in [0, 1]$, of the global maximum peak. In most cases, these minor peaks are not interesting, and many are detected due to the engagement in the adjacent time intervals being zero. This can be done since we know the entire time series a priori. Thirdly, we filter out adjacent peaks – peaks that are located two hours apart can be considered as one the same peak. We define a distance $d_{\text{adj}} \in \mathbb{N}$ and use it to group adjacent peaks. We discard all but the largest peak in any interval where the distance between peaks is no more than d_{adj}.

The hyperparameters ℓ, d, and influence can be varied to adjust the sensitivity of the algorithm. Increasing the lag parameter will lead to a smoother mean and standard deviation, which can be useful if the data is stationary. A lag of about $\ell = 10$ hours is suitable; we do not want to take engagement from too large a period into account when judging whether a given point is a peak. Increasing the threshold d decreases sensitivity of the algorithm - we find that setting $d = 1.5$ gives reasonable results. The influence parameter was set to a value of $w = 0.8$ to let new data points have a relatively large impact on the mean or median.

To start detection at the time when the root tweet was posted, $t = 0$, there needs to be data points prior to the collected time series of engagement. In practice, we concatenate the engagement bin values to a vector of ℓ zeroes. We use $\ell = 10$, $d = 1.5$, $w = 0.8$, $v = 0.15$, and $d_{\text{adj}} = 2$ as input to Algorithm 2.

A.3 Second Order System Solution

Starting from (2.3) and letting $u(t) = \delta(t)$ we have

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix}}_{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \underbrace{\begin{pmatrix} 0 \\ \rho \end{pmatrix}}_{b} \delta(t), \tag{A.1}$$

The general solution to the system is given by [10]

$$\begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} = e^{At} \begin{pmatrix} x_1(0) \\ x_2(0) \end{pmatrix} + \int_0^t e^{A(t-\tau)} b g(\tau) d\tau, \tag{A.2}$$

where

$$e^{At} = \sum_{k=0}^{\infty} \frac{t^k}{k!} A^k. \tag{A.3}$$

We have

$$A^m = \begin{pmatrix} \alpha^m & \beta \sum\limits_{j=0}^{m-1} \alpha^j \gamma^{m-1-j} \\ 0 & \gamma^m \end{pmatrix} \tag{A.4}$$

for $m > 0$, and with $\gamma < \alpha < 0$, we let

$$\Sigma_m = \beta \sum_{j=0}^{m-1} \alpha^j \gamma^{m-1-j} = \beta \gamma^{m-1} \sum_{j=0}^{m-1} \left(\frac{\alpha}{\gamma}\right)^j = \beta \gamma^{m-1} \frac{1 - \left(\frac{\alpha}{\gamma}\right)^m}{1 - \frac{\alpha}{\gamma}} = \beta \frac{\alpha^m - \gamma^m}{\alpha - \gamma}. \tag{A.5}$$

With initial conditions $x_1(t) = x_2(t) = 0$ for $t \leq 0$, we get for $t > 0$:

$$\begin{aligned} \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix} &= \int_0^t \sum_{k=0}^{\infty} \frac{(t-\tau)^k}{k!} A^k \begin{pmatrix} 0 \\ \rho \end{pmatrix} \delta(\tau) d\tau \\ &= \int_0^t \sum_{k=0}^{\infty} \frac{(t-\tau)^k}{k!} \begin{pmatrix} \alpha^k & \Sigma_k \\ 0 & \gamma^k \end{pmatrix} \begin{pmatrix} 0 \\ \rho \end{pmatrix} \delta(\tau) d\tau \\ &= \int_0^t \sum_{k=0}^{\infty} \frac{(t-\tau)^k}{k!} \begin{pmatrix} \rho \Sigma_k \\ \rho \gamma^k \end{pmatrix} \delta(\tau) d\tau \\ &= \rho \sum_{k=0}^{\infty} \frac{t^k}{k!} \begin{pmatrix} \Sigma_k \\ \gamma^k \end{pmatrix} u(t). \end{aligned} \tag{A.6}$$

Now,

$$x_1(t) = \rho \sum_{k=0}^{\infty} \frac{t^k}{k!} \Sigma_k H(t) = \rho\beta \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{\alpha^k - \gamma^k}{\alpha - \gamma} H(t) = \frac{\rho\beta}{\alpha - \gamma} (e^{\alpha t} - e^{\gamma t}) H(t), \tag{A.7}$$

and

$$x_2(t) = \rho \sum_{k=0}^{\infty} \frac{t^k}{k!} \gamma^k H(t) = \rho e^{\gamma t} H(t). \tag{A.8}$$

The sum of the two becomes

$$y(t) = x_1(t) + x_2(t) = \frac{\rho\beta}{\alpha - \gamma} e^{\alpha t} H(t) + \rho\left(1 - \frac{\beta}{\alpha - \gamma}\right) e^{\gamma t} H(t). \tag{A.9}$$

A.4 Twitter API Details

This section contains tables listing the details of the Twitter API. Tables A.3 and A.4 present the attributes of user and tweet objects respectively. Table A.5 lists the domain classifications used by Twitter to annotate tweets. Table A.6 shows a subset of available endpoints of the API, and their rate limits.

Table A.3: *Attributes available for user objects from the Twitter API.*

User field	Explanation
`id`	Unique account ID (appears as `author_id` in Table A.4).
`name`	Name of user.
`username`	Twitter alias used in @-mentions. Modifiable by the user.
`created_at`	Creation date of user account (UTC).
`description`	User provided biography, if available.
`entities`	Information on hashtags, URLs, user mentions in the user description.
`location`	Location specified in user profile.
`pinned_tweet_id`	Tweet ID of the users pinned tweet.
`profile_image_url`	The URL to the user profile picture.
`protected`	Indicates whether the tweets of the user are non-public.
`public_metrics`	Counts on the number of followers, following, tweets, and list presence of the user.
`url`	The URL specified in the user profile, if present.
`verified`	Indicates whether Twitter has verified the identity of the user.
`withheld`	Details on withheld content.

Table A.4: *Attributes available for tweets objects from the Twitter API.*

Tweet field	**Explanation**
`id`	Unique 64-bit Tweet ID.
`text`	UTF-8 text content of a tweet.
`attachments`	Keys to media attachments, such as polls, or images.
`author_id`	Unique user ID of the tweet author.
`context_annotations`	List of objects (domain-entity pairs) that the tweet refers to. Generated by Twitters named entity recognition algorithms. See Table A.5.
`conversation_id`	Tweet ID of the conversation root message.
`created_at`	Time of posting in UTC.
`entities`	Information on hashtags, URLs, mentions, and other annotations in the tweet.
`geo`	Optional geographical tag of tweet.
`in_reply_to_user_id`	User ID of the author of the tweet to which this tweet replies, if applicable.
`lang`	Language of the tweet, if detected by Twitter.
`non_public_metrics`	Hidden engagement metric, such as impression count, the times users clicked on the embedded link or the author profile. Not accessible to anyone except the author.
`organic_metrics`	Engagement metrics of the tweet in an organic context (non-promotional). Not accessible to anyone except the author.
`possibly_sensitive`	Indicates whether the URL contained in the tweet points to sensitive content. Does not relate to the content of the tweet itself.
`promoted_metrics`	Engagement metrics of the tweet when presented in the context of promotion. Not accessible to anyone except the author.
`public_metrics`	Engagement metrics of the tweets: number of retweets, replies, quotes, and likes.
`referenced_tweets`	A (nested) list of referenced tweets. The engagement type (reply or quote) is indicated as well as the tweet ID.
`reply_settings`	Indicated who may reply to the tweet, everyone, mentioned users or followers only.
`source`	Platform from which the tweet was posted, e.g., Twitter Web App.
`withheld`	Indicates, e.g., copyright infringement if present.

Table A.5: *Table of domains used by Twitter to classify tweets. In addition to these, there are at least two meta domains: "Entities [Entity Service]" (30), and "Unified Twitter Taxonomy" (131).*

3 - TV Shows	79 - Video Game Hardware
4 - TV Episodes	84 - Book Music Genre
6 - Sports Events	85 - Book Genre
10 - Person	86 - Movie
11 - Sport	87 - Movie Genre
12 - Sports Team	88 - Political Body
26 - Sports League	89 - Music Album
27- American Football Game	90 - Radio Station
28 - NFL Football Game	91 - Podcast
35 - Politicians	92 - Sports Personality
38 - Political Race	93 - Coach
39 - Basketball Game	94 - Journalist
40 - Sports Series	110 - Viral Accounts
45 - Brand Vertical	114 - Concert
46 - Brand Category	115 - Video Game Conference
47 - Brand	116 - Video Game Tournament
48 - Product	117 - Movie Festival
49 - Product Version	118 - Award Show
54 - Musician	119 - Holiday
55 - Music Genre	120 - Digital Creator
56 - Actor	122 - Fictional Character
58 - Entertainment Personality	123 - Ongoing News Story
60 - Athlete	130 - Multimedia Franchise
65 - Interests and Hobbies Vertical	132 - Song
66 - Interests and Hobbies Category	136 - Video Game Personality
67 - Interests and Hobbies	137 - eSports Team
68 - Hockey Game	138 - eSports Player
71 - Video Game	139 - Fan Community
78 - Video Game Publisher	

Table A.6: A selection of endpoints of the Twitter API v2.

Endpoint	Use	Rate limit, requests/15 min
Recent search	Retrieve conversation thread, retweets, etc.	450
Full archive search	Retrieve conversation thread, retweets, etc.	300
Sampled stream	Sample random tweets	50
Retweets lookup (by user)	Get user ID of retweeters	75
Quote tweets lookup	Retrieve quotes of a tweet	75
Follows lookup	Extract follower/following network topology	15
User lookup	Obtain user profile from ID or name	900
Tweet counts	Get tweet volume of query without it counting towards tweet cap	300
Filtered stream	Stream tweets in real time	50
Timeline	Get tweets from a user's timeline	1500

A.5 Supplementary Graphs

In this section we show a few examples of models of type (2.1) and (2.3) fitted to engagement data of different categories. Figures A.1-A.4 show conversations of type (1), (2), (3) and (4) as defined in Section 2.4.1, respectively, where there is no delay in the inital peak, and A.5-A.8 show the corresponding cases with a delay in the inital peak.

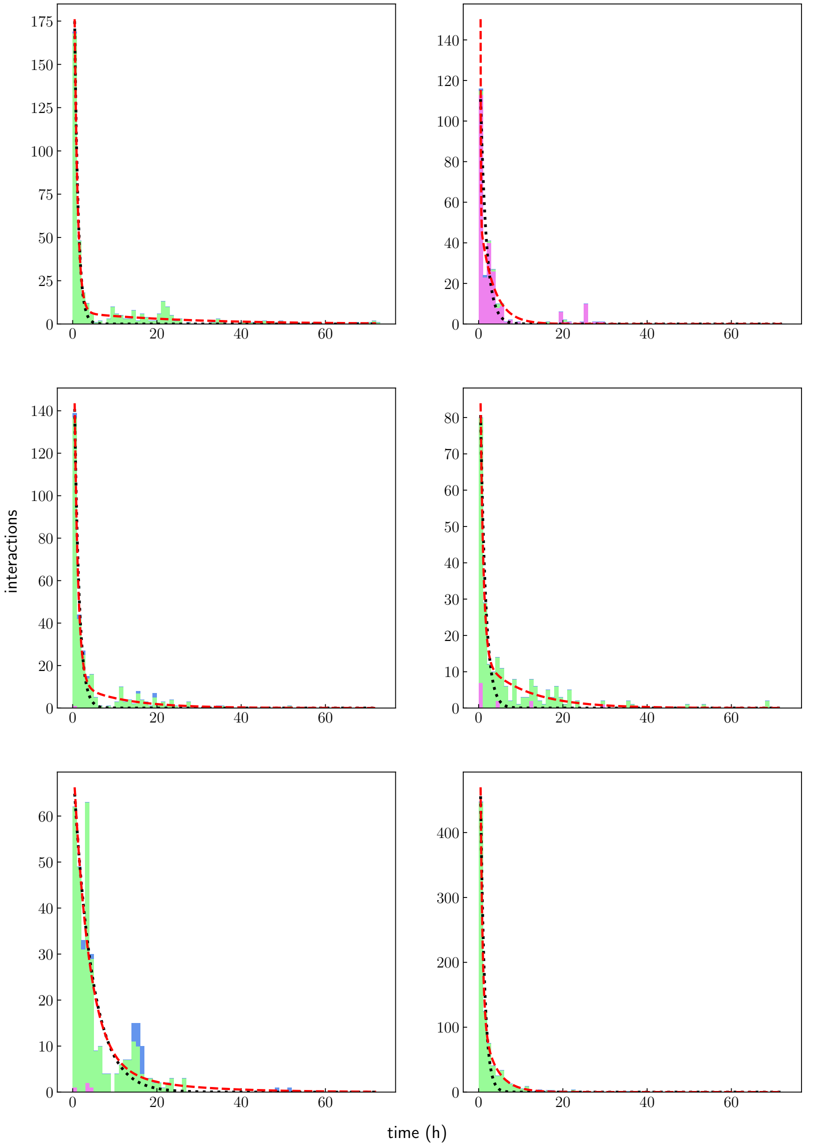


Figure A.1: *Models fitted to conversations of type (1) with a peak during the first hour. Dotted black line: first order system. Dashed red line: second order system.*

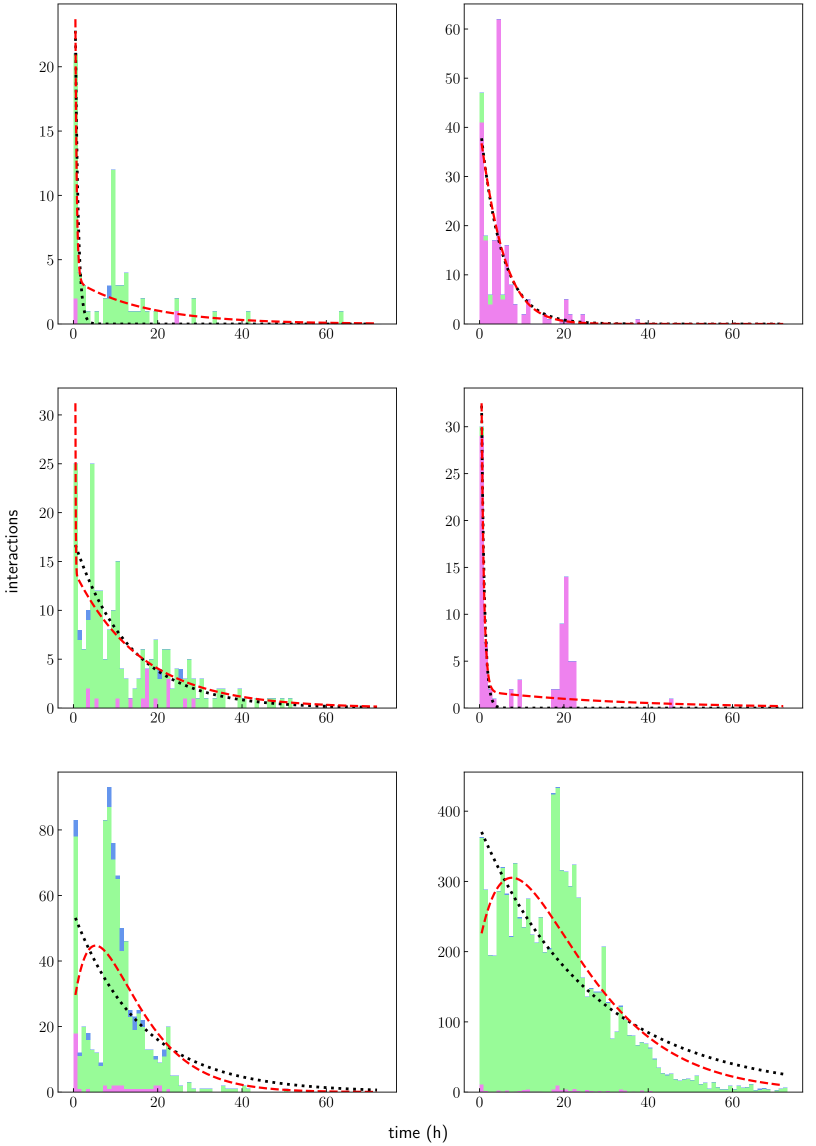


Figure A.2: *Models fitted to conversations of type (2) with a peak during the first hour. Dotted black line: first order system. Dashed red line: second order system.*

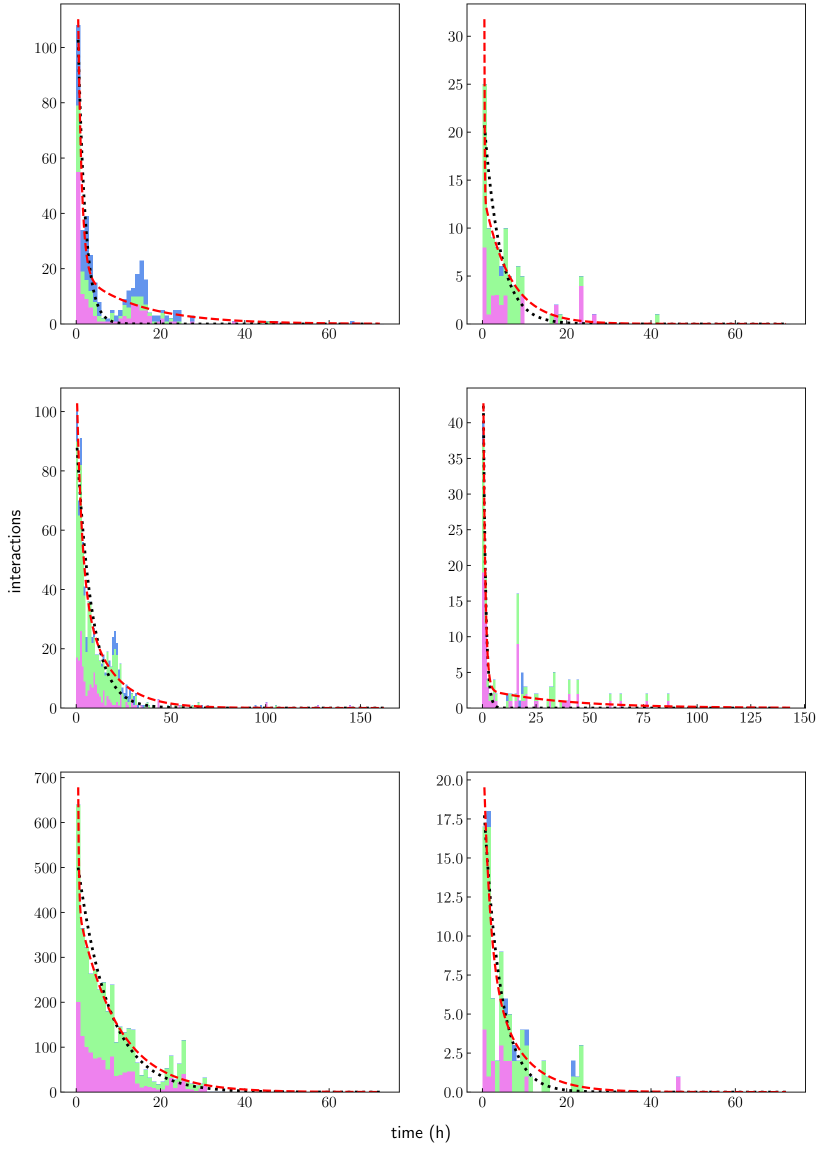


Figure A.3: *Models fitted to conversations of type (3) with a peak during the first hour. Dotted black line: first order system. Dashed red line: second order system.*

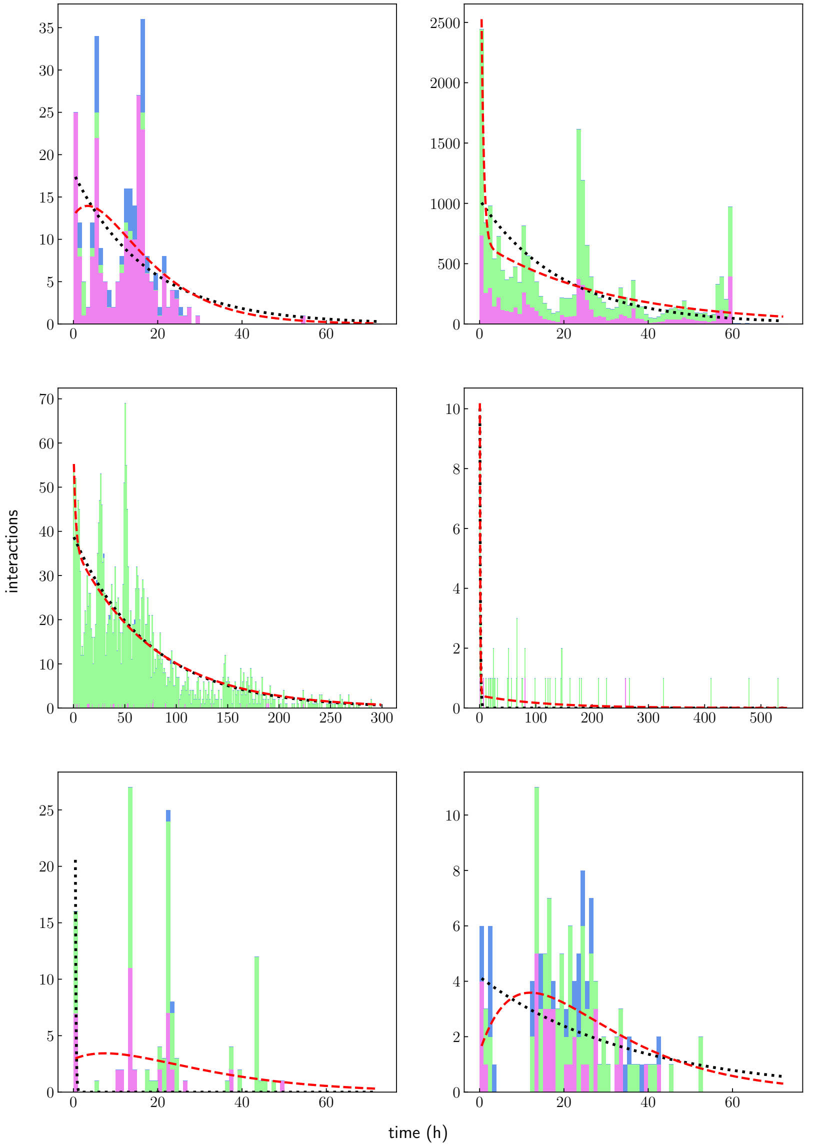


Figure A.4: *Models fitted to conversations of type (4) with a peak during the first hour. Dotted black line: first order system. Dashed red line: second order system.*

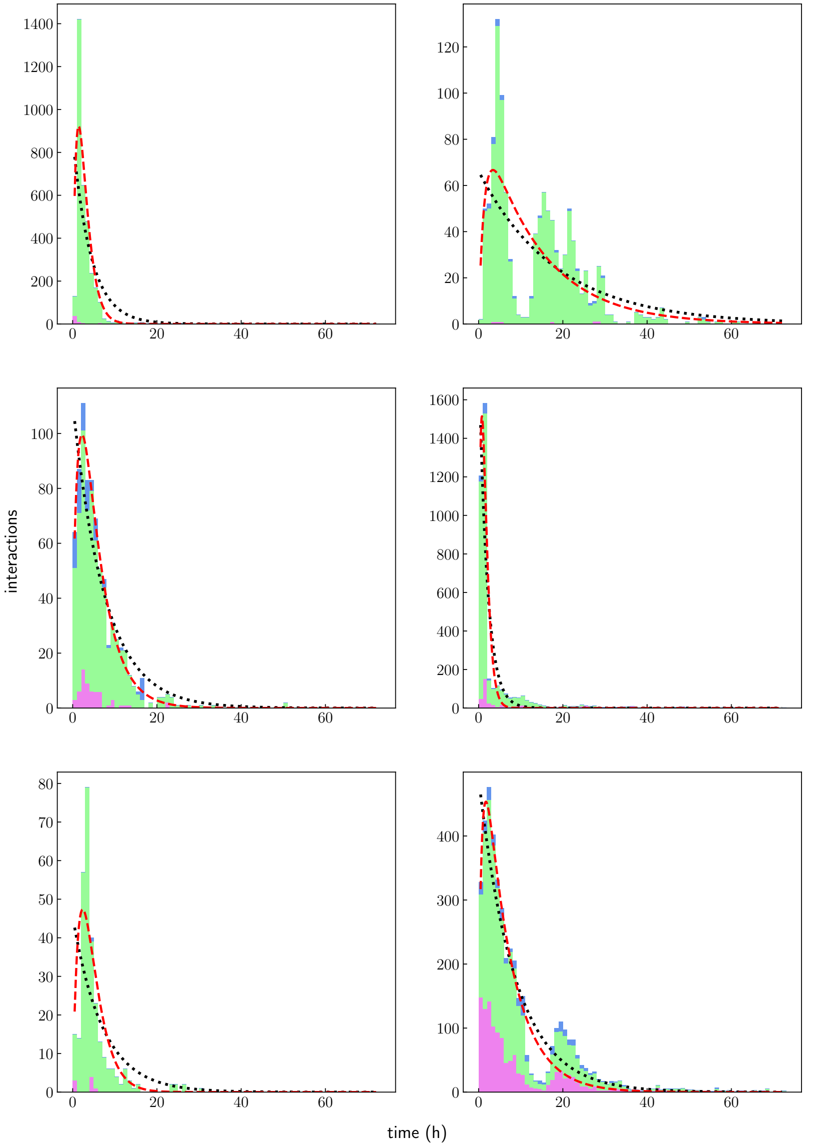


Figure A.5: *Models fitted to conversations of type (1) with a delayed initial peak. Dotted black line: first order system. Dashed red line: second order system.*

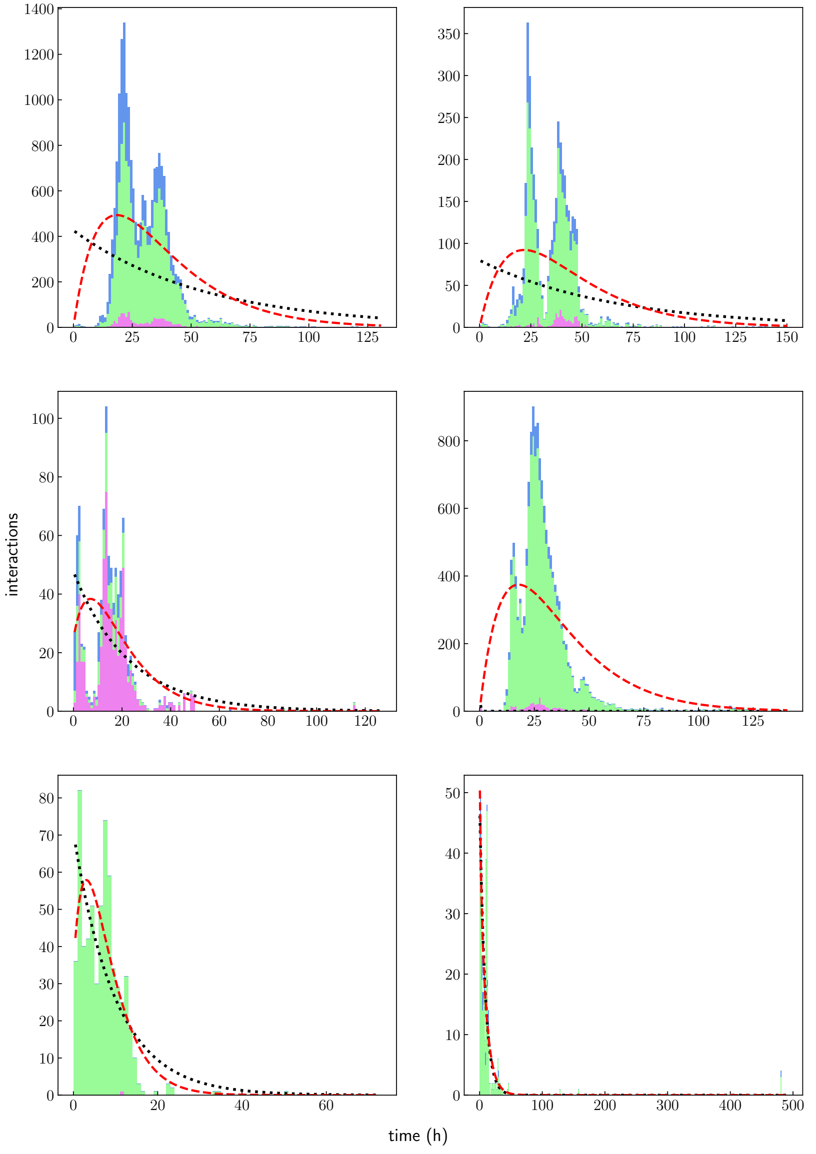


Figure A.6: *Models fitted to conversations of type (2) with a delayed initial peak. Dotted black line: first order system. Dashed red line: second order system.*

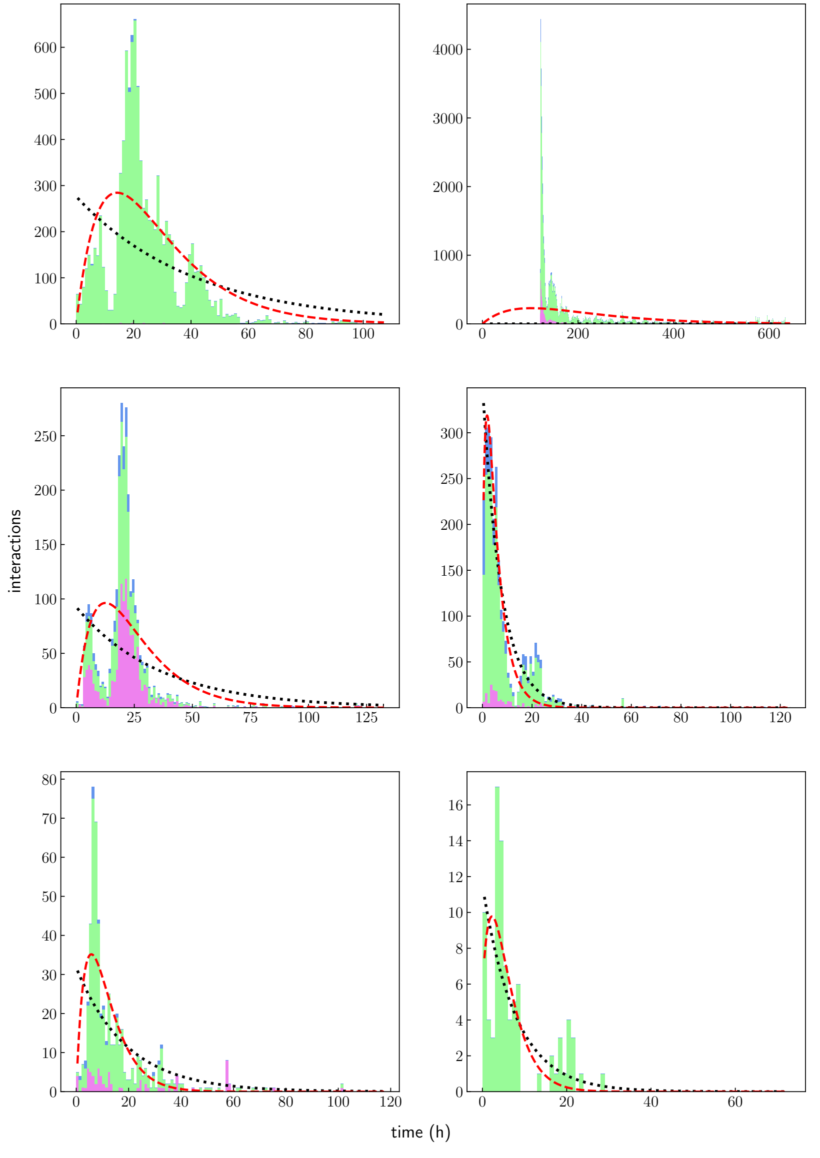


Figure A.7: *Models fitted to conversations of type (3) with a delayed initial peak. Dotted black line: first order system. Dashed red line: second order system.*

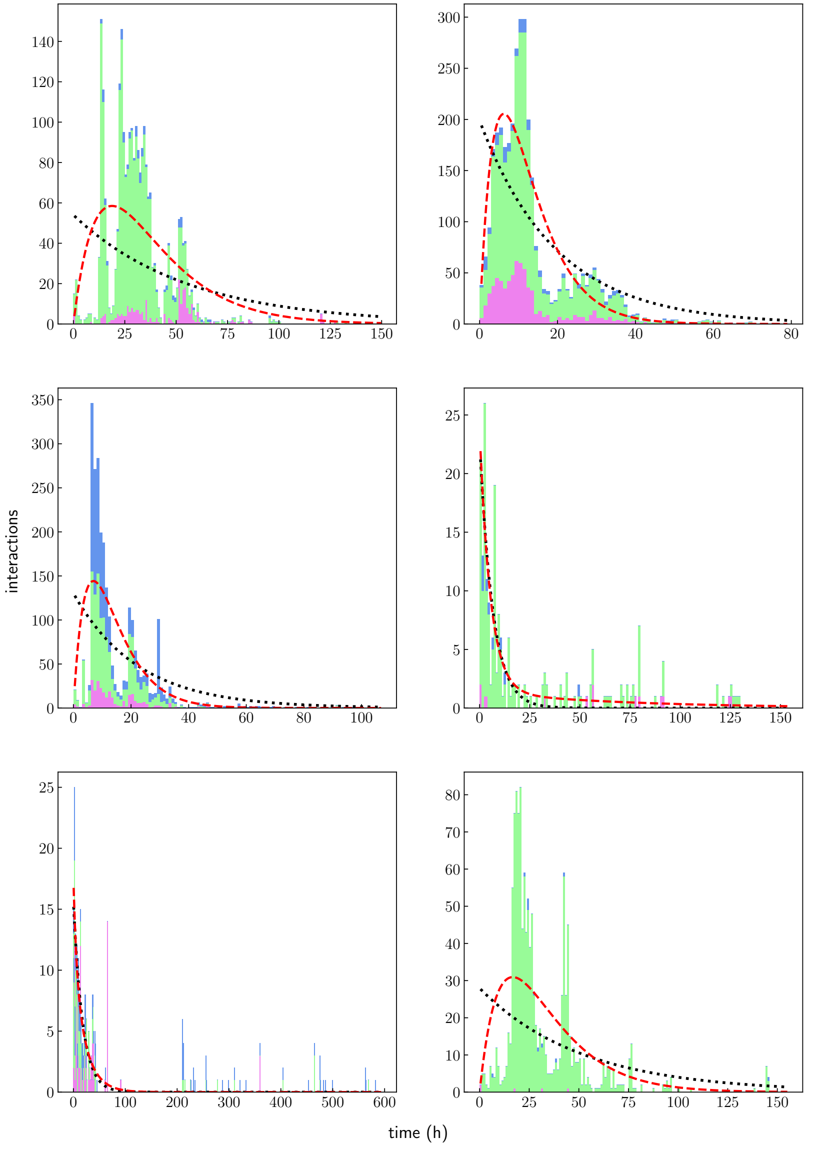


Figure A.8: *Models fitted to conversations of type (4) with a delayed initial peak. Dotted black line: first order system. Dashed red line: second order system.*

Bibliography

[1] Albert-László Barabási. The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039):207–211, 2005.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[3] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge University Press, Cambridge, 2016. ISBN 9781107076266 1107076269.

[4] Cristian Candia, C Jara-Figueroa, Carlos Rodriguez-Sickert, Albert-László Barabási, and César A Hidalgo. The universal decay of collective memory and attention. *Nature Human Behaviour*, 3(1):82–91, 2019.

[5] Meeyoung Cha, Hamed Haddadi, Fabricio Benevenuto, and Krishna Gummadi. Measuring user influence in twitter: The million follower fallacy. In *Proceedings of the international AAAI conference on web and social media*, volume 4, pages 10–17, 2010.

[6] Randy Harris. *Modern physics*. Pearson/Addison Wesley, 2008.

[7] Ravi Kumar, Mohammad Mahdian, and Mary McGlohon. Dynamics of conversations. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 553–562, 2010.

[8] Vito Latora, Vincenzo Nicosia, and Giovanni Russo. *Complex networks: principles, methods and applications*. Cambridge University Press, 2017.

[9] Eugene Leypunskiy, Emre Kıcıman, Mili Shah, Olivia J. Walch, Andrey Rzhetsky, Aaron R. Dinner, and Michael J. Rust. Geographically resolved rhythms in twitter use reveal social pressures on daily activity patterns. *Current Biology*, 28(23):3763–3775.e5, 2018. ISSN 0960-9822.

[10] Lennart Ljung and Torkel Glad. *Modellbygge och simulering*. Studentlitteratur, Lund, 2.6 edition, 2004. ISBN 9144024436.

[11] Philipp Lorenz-Spreen, Bjarke Mørch Mønsted, Philipp Hövel, and Sune Lehmann. Accelerating dynamics of collective attention. *Nature communications*, 10(1):1–9, 2019.

[12] Seth A Myers and Jure Leskovec. The bursty dynamics of the twitter information network. In *Proceedings of the 23rd international conference on World wide web*, pages 913–924, 2014.

[13] Ernest Rutherford and Frederick Soddy. Lx. radioactive change. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 5(29):576–591, 1903.

[14] Eleni Stai, Eirini Milaiou, Vasileios Karyotis, and Symeon Papavassiliou. Temporal dynamics of information diffusion in twitter: Modeling and experimentation. *IEEE Transactions on Computational Social Systems*, 5(1): 256–264, 2018.

[15] Io Taxidou and Peter M Fischer. Online analysis of information diffusion in twitter. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 1313–1318, 2014.

[16] Jean-Paul van Brakel. Robust peak detection algorithm using z-scores, 2014. URL `https://stackoverflow.com/questions/22583391/peak-signal-detection-in-realtime-timeseries-data/22640362#22640362`.

[17] Frederik Wolf, Philipp Lorenz-Spreen, and Sune Lehmann. Successive cohorts of twitter users show increasing activity and shrinking content horizons. *arXiv preprint arXiv:2108.08641*, 2021.

[18] Fang Wu and Bernardo A Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599–17601, 2007.

[19] Qingyuan Zhao, Murat A Erdogdu, Hera Y He, Anand Rajaraman, and Jure Leskovec. Seismic: A self-exciting point process model for predicting tweet popularity. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1513–1522, 2015.